

IMAGE SEARCH METHOD AND IMAGE SEARCH ENGINE DEVICE

Publication number: JP2002132832 (A)

Publication date: 2002-05-10

Inventor(s): LEE JAIRES; STEPHEN ROBERT LAWRENCE

Applicant(s): NIPPON ELECTRIC CO

Classification:

- international: **G06F17/30; G06F3/00; G06F3/048; G06F17/30; G06F3/00; G06F3/048;** (IPC1-7): G06F17/30; G06F3/00


- European: G06F17/30W1

Application number: JP20010237654 20010806

Priority number(s): US19970062958P 19971010; US19980113751 19980710

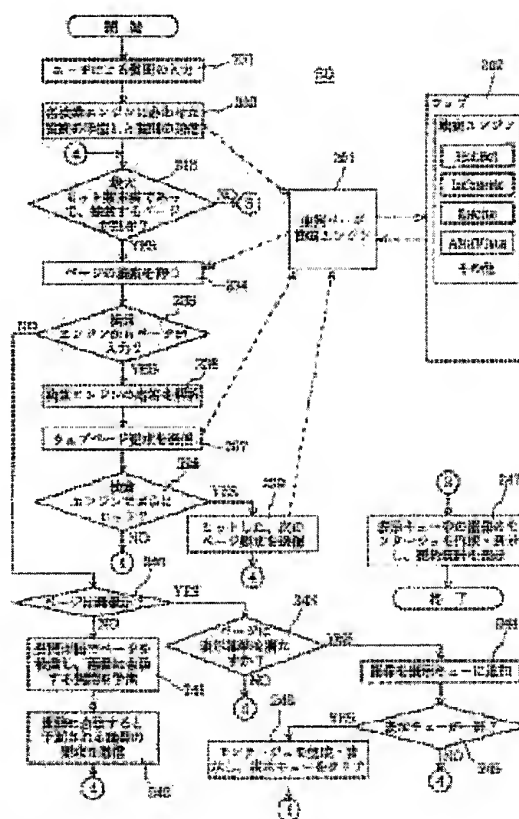
Also published as:

 US6999959 (B1)

 JP11191114 (A)

Abstract of JP 2002132832 (A)

PROBLEM TO BE SOLVED: To speedily and accurately present an image that a user really requires to the user when information on the Internet is searched for. **SOLUTION:** The image search engine and search method send a question to multiple third-party search engines on the Internet (web 202) (step 232) and analyzes answers from the respective search engines (step 236) to add images which are considered to match the question to a display queue (step 244). When the display queue becomes full, montages of images are generated and displayed (step 246).



Data supplied from the **esp@cenet** database — Worldwide

Reference 4

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-132832

(P2002-132832A)

(43) 公開日 平成14年5月10日 (2002.5.10)

(51) Int.Cl. ⁷	識別記号	F I	テ-マ-ド [*] (参考)
G 0 6 F 17/30	3 8 0	G 0 6 F 17/30	3 8 0 F 5 B 0 7 5
	1 1 0		1 1 0 F 5 E 5 0 1
	1 7 0		1 7 0 B
	3 5 0		3 5 0 C
3/00	6 5 1	3/00	6 5 1 A

審査請求 有 請求項の数23 O L (全 36 頁) 最終頁に続く

(21) 出願番号 特願2001-237654(P2001-237654)

(62) 分割の表示 特願平10-286599の分割

(22) 出願日 平成10年10月8日 (1998.10.8)

(31) 優先権主張番号 60/062958

(32) 優先日 平成9年10月10日 (1997.10.10)

(33) 優先権主張国 米国 (US)

(31) 優先権主張番号 09/113751

(32) 優先日 平成10年7月10日 (1998.7.10)

(33) 優先権主張国 米国 (US)

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 リー ジャイレス

アメリカ合衆国、 ニュージャージー

08540、 プリンストン、 インディペン

デンス ウェイ 4 エヌ・イー・シー・

リサーチ・インスティテュート・インク

内

(74) 代理人 100088328

弁理士 金田 暢之 (外2名)

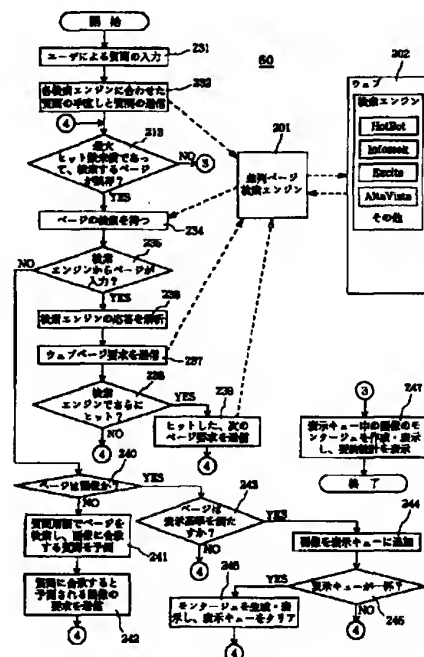
最終頁に続く

(54) 【発明の名称】 画像検索方法及び画像検索エンジン装置

(57) 【要約】

【課題】 インターネット上にある情報を検索する際に、利用者が真に要求している画像を迅速かつ的確にその利用者に提示できるようにする。

【解決手段】 本発明の画像検索エンジンおよび検索方法では、質問をインターネット (ウェブ202) 上にある複数のサードパーティ検索エンジンに送り (ステップ232)、各検索エンジンからの応答を解析することで (ステップ236)、質問に合致すると考えられる画像を表示キューに追加する (ステップ244)。表示キューが一杯になったら、画像のモンタージュを作成して、表示する (ステップ246)。



【特許請求の範囲】

【請求項 1】 コンピュータで実行されキーワードに基づいて画像を検索する画像検索方法であって、質問を複数のサードパーティ画像検索エンジンに送る段階と、

前記サードパーティ画像検索エンジンからの応答を解析することで、前記質問に合致する画像に関する情報を抽出する段階と、

前記質問に合致する画像をダウンロードする段階と、

前記画像のサムネイル画像をユーザーに表示する段階と、を有する画像検索方法。

【請求項 2】 コンピュータで実行されキーワードに基づいて画像を検索する画像検索方法であって、質問を複数のサードパーティテキスト検索エンジンに送る段階と、

前記サードパーティテキスト検索エンジンからの応答を解析することで、前記質問に合致するドキュメントに関する情報を抽出する段階と、

前記質問に合致するドキュメントをダウンロードする段階と、

前記ドキュメントを解析し、前記質問に用いた用語である質問用語と画像タグまたは参照語(reference)との近接度に基づいて、ユーザーの質問と合致しうる画像の配備位置を決定する段階と、

前記画像をダウンロードする段階と、

前記画像のサムネイル画像をユーザーに表示する段階と、を有する画像検索方法。

【請求項 3】 ユーザーによって選択可能であり、画像のサイズ、色または意味論的属性に基づく、画像のフィルター処理を行う段階をさらに有する請求項 1 または 2 に記載の画像検索方法。

【請求項 4】 画像やドキュメントのセクションを分離するのに使用される水平方向バーなどの、ウェブで一般的に使用される画像を同定してフィルター処理する段階をさらに有する請求項 1 または 2 に記載の画像検索方法。

【請求項 5】 類似する画像を同定してフィルター処理を行う段階をさらに有する請求項 1 または 12 に記載の画像検索方法。

【請求項 6】 写真、線画、ロゴ、地図、漫画、ポートレート、ボタン、チャート及び天文画などの画像の種類を同定する段階と、その画像の種類に基づいてフィルター処理を行う段階と、をさらに有する請求項 1 または 2 に記載の画像検索方法。

【請求項 7】 質問に合致する画像を保存して、質問を繰り返すことができるようにする段階と、新規画像のみを提示する段階とをさらに有する請求項 1 または 2 に記載の画像検索方法。

【請求項 8】 画像の種類などのメタ情報を保存することで、新たな質問に対して画像の再ダウンロードを行う

ことなく、前記メタ情報を用いて画像のフィルター処理を行うことができるようにする段階をさらに有する請求項 1 または 2 に記載の画像検索方法。

【請求項 9】 画像全体を表示するとともに可能であればその画像を参照するドキュメントを表示し、ドキュメント中での質問用語を強調表示する段階をさらに有する請求項 1 または 2 に記載の画像検索方法。

【請求項 10】 コンピュータに実装されキーワードに基づいて画像を検索する画像検索エンジン装置であって、

質問を複数のサードパーティ画像検索エンジンに送る手段と、

前記サードパーティ画像検索エンジンからの応答を解析することで、前記質問に合致する画像に関する情報を抽出する手段と、

前記質問に合致する画像をダウンロードする手段と、

前記画像のサムネイル画像をユーザーに表示する手段と、

を有する画像検索エンジン装置。

【請求項 11】 コンピュータに実装されキーワードに基づいて画像を検索する画像検索エンジン装置であって、

質問を複数のサードパーティテキスト検索エンジンに送る手段と、

前記サードパーティテキスト検索エンジンからの応答を解析することで、前記質問に合致するドキュメントに関する情報を抽出する手段と、

前記質問に合致するドキュメントをダウンロードする手段と、

前記ドキュメントを解析し、前記質問に用いた用語である質問用語と画像タグまたは参照語(reference)との近接度に基づいて、ユーザーの質問と合致しうる画像の配備位置を決定する手段と、

前記画像をダウンロードする手段と、

前記画像のサムネイル画像をユーザーに表示する手段と、

を有する画像検索エンジン装置。

【請求項 12】 画像のサイズ、色または意味論的属性に基づいて、選択可能なフィルター処理を画像に行う手段をさらに有する請求項 10 または 11 に記載の画像検索エンジン装置。

【請求項 13】 画像やドキュメントのセクションを分離するのに使用される水平方向バーなどの、ウェブで一般的に使用される画像を同定してフィルター処理する手段をさらに有する請求項 10 または 11 に記載の画像検索エンジン装置。

【請求項 14】 類似する画像を同定してフィルター処理を行う手段をさらに有する請求項 10 または 11 に記載の画像検索エンジン装置。

【請求項 15】 写真、線画、ロゴ、地図、漫画、ポートレート、ボタン、チャート及び天文画などの画像の種類を同定し、その画像の種類に基づいてフィルター処理

を行う手段をさらに有する請求項10または11に記載の画像検索エンジン装置。

【請求項16】 質問に合致する画像を保存して質問を繰り返すことができるようにするとともに、新規画像のみを提示する手段をさらに有する請求項10または11に記載の画像検索エンジン装置。

【請求項17】 画像の種類などのメタ情報を保存することで、新たな質問に対して画像の再ダウンロードを行うことなく、前記メタ情報を用いて画像のフィルター処理を行うことができるようにする手段をさらに有する請求項10または11に記載の画像検索エンジン装置。

【請求項18】 画像全体を表示するとともに可能であればその画像を参照するドキュメントを表示する手段と、ドキュメント中での質問用語を強調表示する手段とをさらに有する請求項10または11に記載の画像検索エンジン装置。

【請求項19】 コンピュータによって実行され、サードパーティの検索エンジンの相対的検索範囲を推定する方法であって、
2個のサードパーティの検索エンジンに対して1群の質問を送る段階と、
各検索エンジンからの結果の完全リストを検索する段階と、
各検索エンジンによってリストアップされた全ページのテキストを検索する段階と、
利用できないページ及び質問と合致しないページをフィルター処理によって除去する段階と、
各検索エンジンからのページのうち残ったページの数と比較する段階と、を有する相対的検索範囲推定方法。

【請求項20】 コンピュータで実行される情報検索方法であって、
疑問の形態の質問を認識する段階と、
前記疑問に対する回答が表現されると考えられる1以上の具体的表現形式からなる集合に前記質問を変換する段階と、
前記変換された質問について検索を行う段階と、を有する情報検索方法。

【請求項21】 各種類の疑問についての前記具体的表現形式がマニュアル操作で記述されている請求項20記載の情報検索方法。

【請求項22】 質問を複数のサードパーティ検索エンジンに送る段階と、
前記サードパーティ検索エンジンからの応答を解析することで、前記質問に合致するドキュメントに関する情報を抽出する段階と、
前記質問に合致するドキュメントの全テキストをダウンロードする段階と、
前記ドキュメント中で、前記質問に用いた用語である質問用語の位置を見つけ、前記質問用語の周囲のテキストを抽出する段階と、

前記質問用語の周囲のテキストを表示する段階と、
前記文脈の共通形式を同定する段階と、を有する検索方法によって、ユーザーが選択するドキュメント中の質問用語の文脈を解析することで、各種類の疑問についての前記具体的表現形式が学習される請求項42記載の情報検索方法。

【請求項23】 コンピュータで実行される質問拡張方法であって、
質問に用いた用語である質問用語を語幹化(stemming)する段階と、
前記質問に対する結果のページの集合において、前記質問用語の共通に出現する変形語(morphological variant)を検索する段階と、
前記共通に出現する変形語を用いて、質問の拡張を行う段階と、を有する質問拡張方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、インターネット(Internet)などのネットワーク上でドキュメント(文書)や画像データ、各種コンテンツ類などを検索する検索方法及び検索エンジンに関し、特に、サードパーティ(第三者)の検索エンジンを用いて検索を実行するメタ検索に関し、特に、画像検索方法及び画像検索エンジンに関する。

【0002】

【従来の技術】多くの有用かつ一般的な検索エンジンが、インターネット上のWWW(WorldWide Web:ワールド・ワイド・ウェブ)の全文(フルテキスト)索引を維持するよう努めている。例えば、現在、インターネットでは、AltaVista(アルタビスタ)、Excite(エキサイト)、HotBot(ホットボット)、Infoseek(インフォシーク)、Lycos(ライコス)及びNorthern Light(ノーザンライト)などの検索サービスから検索エンジンが利用可能である。しかしながら、ウェブの検索は、未だに、遅くて冗長な作業になる場合がある。

【0003】検索サービスに限界があることから、メタ検索エンジンが導入されるようになった。メタ検索エンジンは、AltaVistaやInfoseekなどの複数の検索エンジンに要求を行うことで、ウェブの検索を行うものである。現在のメタ検索エンジンの主たる利点は、複数の検索エンジンの結果を組み合わせることができることと、これらエンジンを検索するのに一貫したユーザーインターフェースを提供できることである。実験結果から、主要な検索エンジンが比較的少ない量のウェブに対してのみインデックス(索引)付けを行い、したがって複数のエンジンの検索結果を組み合わせることによって、他の方法では見つからないと考えられる多くのドキュメントを回答として与えることができることが、明らかになっている。

【0004】現在、多くのメタ検索エンジンが利用可能

である。最も一般的なものをいくつか挙げると、MetaCrawler (メタクラウラー)、Inference Find (インフェレンスファインド)、SavvySearch (サビーサーチ)、Fusion (フュージョン)、ProFusion (プロフュージョン)、Highway 61 (ハイウェイ 61)、Mamma (マンマ)、Quarterdeck WebCompass (クォーターデックウェブコンパス)、Symantec Internet FastFind (シマンテックインターネットファストファインド) および Front WebSeeker (フォアフロントウェブシーカー) などがある。

【0005】

【発明が解決しようとする課題】従来の主要なウェブ検索エンジンにおいては、精度が欠如し、検索範囲が限られており、利用性が限られており、ユーザーインターフェースに限界があり、データベースが古いという問題点があり、これこそが、本発明においてメタ検索エンジンの基本的なテキストのメタ検索能力を向上させようとする主要な動機である。より具体的には、従来の主要なウェブ検索エンジンでは、ウェブが多様であり、ウェブ検索エンジンが比較的簡単な質問を非常に迅速に取り扱うことに注力していることから、多くの場合、精度を欠く検索結果を生じることになる。さらに、「検索エンジンスパミング (spamming)」の実行が一般的となり、それによってユーザーは、検索結果でのページの順位付けを変えるために、そのページに無関係のキーワードを加える可能性がある。検索でヒットした特定のページと質問との関連性は、多くの場合、ページのダウンロードを待って、そのページで質問用語を見出した後に初めて明らかになる。

【0006】各種の検索エンジンを用いた経験から、個々の検索エンジンの検索範囲が比較的小さいことが示唆される。すなわち、2 つ目のエンジンを用いて検索を行うと、最初のエンジンによっては検索結果として与えられなかったドキュメントが、いくつか見つかる場合が多いと考えられる。検索エンジン AltaVista がドメインごとにインデックス付けされるページ数を制限していることや、各検索エンジンが、それぞれ、インデックス付けするページの選択に異なる戦略を持っていることが、示唆されている。実験結果からは、ある一つの検索エンジンの検索範囲が非常に限られていることが確認されて

いる。

【0007】さらに、検索エンジン及び／またはネットワークにおける問題のために、最も迅速に反応するエンジンは、時間を経るにつれて変わる。検索エンジンの有用性を高める多くの特徴を付与することが可能である。集中化された検索エンジンデータベースは、常に内容が古いものである。新たな情報が利用できるようになると、それに索引付けする時との間には、時間的遅れがある。

【0008】本発明の目的は、メタ検索エンジンを改良

することにある。

【0009】本発明の別の目的は、各ドキュメントを分析し、質問用語の周囲のローカルな文脈を表示するメタ検索エンジンを提供することにある。

【0010】本発明のさらに別の目的は、既存の検索方法よりも効率が向上した検索方法を提供することにある。

【0011】本発明のさらに別の目的は、質問用語の文脈の表示、より進んだ重複検出、結果のプログレッシブな表示 (データが入力し次第、順次表示する表示方法)、ビュー (表示) 時のページにおける質問用語の強調 (ハイライト表示)、大きいページで質問用語を見つけるためのクイック・ジャンプ・リンク (quick jump link) の挿入、固有の表現形式を用いることによるある種の質問に関する精度の大幅な改善、関連性の順位付けの改善、クラスタリングの改善および画像検索を行うことができるメタ検索エンジンを提供することにある。

【0012】

【課題を解決するための手段】上述した目的および他の目的は、コンピュータで実行されるメタ検索エンジンおよび検索方法によって達成される。本発明の方法によれば、多くのサードパーティ (第三者) の検索エンジンに質問を送り、それらサードパーティの検索エンジンからの応答を解析することで、その質問に合致するドキュメントに関する情報を抽出する。質問に合致するドキュメントの全テキストをダウンロードし、ドキュメント中の質問用語 (質問に用いた用語) の位置を見つけ出す。質問用語の周囲のテキストを抽出して、そのテキストを表示する。

【0013】本発明のメタ検索エンジンは、検索でのヒットに対応する実際のページをダウンロードし、質問用語についてそれらページ中での検索を行う。検索エンジンは、そのページの概要ではなく、その質問用語が出てくる文脈 (コンテキスト) を提供する (現在利用できる検索エンジンおよびメタ検索エンジンで、このような機能を提供するものはない)。質問用語が出てくる文脈は、典型的には、他の検索エンジンが使用する概要や要約より、質問とページとの関連性をかなり良好に示すものであり、必要な情報が含まれていないことを知ることのみを目的としてページを見るのを回避する上で、役立つ場合が多い。その文脈は、必要とする文脈とは異なる文脈で生じ得る用語を検索が含む場合は常に、特に役立つものとなろう。ユーザは、ユーザー質問用語の前後いずれかの側での文字数によって、文脈の量を指定する。英数字以外の文字のほとんどは、その文脈からフィルタ処理によって除去されて、より読みやすく、より有益な結果を提供する。

【0014】結果は、全てのページをダウンロードした後ではなく、個々のページのダウンロードおよび解析を行った後にその都度もたらされる。最初の結果は、検索

エンジンの平均応答時間より速く表示されるのが普通である。最も速いサイトを最初に解析・表示するので、複数ページによって必要な情報が提供される場合は、メタ検索エンジンこのアーキテクチャーは、有効なものとなり得る。

【0015】検索でのヒットに対応する全ページのビューを行う場合、それらのページを質問用語を強調表示するためにフィルタ処理され、また、各質問用語がそれぞれ最初に出現した箇所にジャンプするためのリンクが、そのページのトップ（先頭部分）に挿入される。質問用語の各出現箇所でのリンクは、それぞれの用語の次の出現箇所にジャンプするものである。質問用語の強調表示は、質問用語とページの関連性を迅速に確認する上で役立つ。リンクは、大きいドキュメント中で質問用語を迅速に見つける上で役立つ。

【0016】すでに利用できないページを確認することができる。そのようなページは、応答の最後にリストアップされる。他のメタ検索サービスの中には「無効リンク」検出を行うものもあるが、その機能はデフォルトではオフになっているのが普通であり、全ページをチェックするまで結果が返ってこない。しかしながら、漸増的かつ迅速に結果を与えることができる本発明のメタ検索エンジンの場合、この機能は検索エンジンのアーキテクチャーに本来備わっているものである。

【0017】検索用語をもちや含まないページ、または質問と適切に合致しないページを確認することができる。そのようなページは、質問と適切に合致するページの後にリストアップされる。それは非常に重要なものとなり得る。というのは、異なるエンジンは異なる関連付け技術を用いており、1個のエンジンが適合性に乏しい結果を返す場合、それによって標準的なメタ検索法から得られる結果は不十分なものとなり得るからである。

【0018】追加ヒット要求という冗長なプロセスを回避することができる。本発明のメタ検索エンジンは、個々の検索エンジン応答から、検索にヒットする次のページを要求するためのURL (Uniform Resource Locator) をどのように抽出するかを理解している。より進んだ重複ページ検出が行なわれる。関連する文脈文字列が同一である場合、ページは重複していると考えられる。それによって、ページが異なるヘッダまたはフッタを有する場合に、重複を検出することができる。

【0019】Kirschの米国特許5659732号明細書には、メタ検索法による関連性の順位付け方法であって、基礎となる検索エンジンに変更を加えて、ドキュメント中での各検索用語の出現回数およびデータベース全体での出現回数などの追加情報を取得する方法が開示されている。本発明のメタ検索エンジンの場合、実際のページをダウンロードおよび解析することから、そのような方法は必要ない。従って、本発明では、各種の検索エンジンが検索結果として与えたドキュメントに対し

て、一定の順位付け尺度を適用することが可能である。現在、本発明の検索エンジンでは、ドキュメント中に出てくる質問用語の数が少なくなる順にページを表示する（最初の数ページに質問用語全てを含むものがない場合、検索エンジンは、最初に、それまでに一つのページ中で見つかった最大質問用語数を含む結果を表示する）。全てのページをダウンロードしたのち検索エンジンは、簡単な関連性尺度にしたがってページの再リストアップする。

【0020】その尺度は、下記式に示すように、現在、ドキュメント中にある質問用語の数、質問用語間の近さおよび用語の頻度に基づくものである。なお、通常の場合、ドキュメント中頻度の逆数もまた有用である (Salt on, G., "Automatic text processing: the transformation, analysis and retrieval of information by computer," Addison-Wesley (1989))。

【0021】

【数1】

$$R = c_1 N_p + \left(c_2 - \frac{\sum_{i=1}^{N_p-1} \sum_{j=i+1}^{N_p} \min(d(i,j), c_2)}{\sum_{k=1}^{N_p-1} (N_p - k)} \right) \frac{c_2 + N_e}{c_1 + c_3}$$

式中、 N_p はドキュメント中にある質問用語の数であり（各用語はそれぞれ1回だけ数える）、 N_e はドキュメント中の質問用語の総数であり、 $d(i, j)$ はドキュメント中に存在するi番目とj番目の質問用語の間の最短距離であり（現在は、文字数によって示している）、 c_1 は関連性尺度Rの全体的大きさを制御する定数であり、 c_2 は有用と考えられる質問用語間の最大距離を指定する定数であり、 c_3 は用語頻度の重要性を指定する定数である。現在、 $c_1 = 100$ 、 $c_2 = 5000$ 、 $c_3 = 10c_1$ である。この尺度は、複数の質問用語を含むページに使用される。見つかった質問用語が1個のみである場合は、ページの最初からの用語の距離を用いる。

【0022】この順位付け基準は、ウェブ検索で特に有用である。ウェブ上の複数の用語についての質問によって、全ての用語を含むドキュメントが結果として得られる場合が多いが、それらの用語は、ドキュメント中では非常に離れていて、ページの無関係な部分にある場合がある。例えば、単一のウェブページで上にアーカイブ（保管）されている別個のUsenet（インターネット上のニュースグループ）メッセージ、あるいはブックマークリストを含むページでの別個のブックマークなどの場合である。

【0023】本発明の検索エンジンは、検索シンタックスに関して、最小公分母法を使わない。この検索エンジンは、ブール代数シンタックスなどの、全ての一般的な検索フォーマットをサポートする。個々の質問シンタックスに合致させるために、質問は動的に変化させられる。この検索エンジンは、質問結果を追跡し、所与の質問に合致する新たなドキュメントが見つかった場合に、

その旨をユーザに自動的に知らせることができる。この検索エンジンは、所定のページのテキストを追跡して、テキストに変更があった場合に、どの行が変更されたかをユーザに知らせることができる。この検索エンジンには、既存の検索エンジンで行われるクラスタリングに対して改良されたクラスタリング技術を含むものである。具体的な表現形式検索法は、ある種の質問についての検索精度を飛躍的に高め得るものである。新たな質問拡張法は、自動的にインテリジェントな質問拡張を行うことができる。

【0024】本発明のメタ検索エンジンに容易に付与できると考えられる別の機能には、改善された関連性尺度、別の順位付け法（例：サイトごとの順位付け）、フィールド検索（例：ページのタイトル、Usenetメッセージの題名、ハイパーリンクテキスト）、特定のサーチエンジンに質問をルーティングするための規則及び／または学習方法、単語の意味の明確化、関連性のフィードバックなどがある。

【0025】

【発明の実施の形態】次に、本発明の好ましい実施の形態について、図面を参照して説明する。

【0026】本発明のメタ検索エンジンの基本的特徴の一つは、各ドキュメントを解析し、質問用語の周囲のローカル文脈を表示するというものである。ドキュメントの要約や概要ではなく、ローカル文脈を表示することの利点は、ドキュメントがユーザーの具体的な質問に答えるものであるか否かを、そのユーザーがより容易に判断できるというものである。本質的にこの方法は、コンピュータが特定のドキュメントの関連性を正確に決定できないことを認め、そのような能力に代わって、ユーザーが関連性を迅速に判断する上で最良の形で情報を配列するものである。したがってユーザーは、質問用語のローカル文脈を迅速にスキャンすることで、関連性の高いドキュメントを見いだすことができる。この方法は簡潔であるが、特にデータベースが非常に大きく、多様で、しかも系統化が不十分なウェブ検索の場合に、非常に有効なものとなり得る。

【0027】複数のデータベースへの質問およびそのデータベースからの結果の照合の考え方は新しいものではない。PLS、Lexis-NexisおよびVerityなどの企業は、複数の異種のデータベースの結果を統合するシステムをかなり以前から作っている。一般的かつ有用なMetaCrawler（メタクラウラー）サービスのよう、他の多くのウェブメタ検索サービスがある。MetaCrawlerと同様のサービスには、SavvySearch、Inference Find、Fusion、ProFusion、Highway 61、Mamma、Quarterdeck Web Compass、Metabot、Symantec Internet FastFindおよびWebSeekerなどがある。

【0028】図1は、本発明の好ましい実施の形態のメタ検索エンジンのホームページ（ここに示す例は、NE

CI (NEC Research Institute Inc.)メタ検索エンジンのホームページである)を示している。一番上のバー12には、オプションページやヘルプページにジャンプし、あるいは「意見と問題報告」を送信するためのリンクがある。質問は、「検索 (Find)」ボックス14に入力する。どの検索エンジンを検索のために使用するかの選択は、その下の行にある適切な選択肢をクリックすることで行う。この場合のオプションは、現在のところ、以下の通りである。

10 【0029】1. ウェブ標準的なウェブ検索エンジン：(a)AltaVista、(b)Excite、(c)Infoseek、(d)HotBot、(e)Lycos、(f)Northern Light、(g)WebCrawler及び(h)Yahoo

2. Usenetデータベース—UseNetニュースグループのインデックス：(a)AltaVista、(b)DejaNews、(c)Reference.com

3. 報道—新聞記事およびオンラインニュースサービス（ニュースワイヤー：news wire）：(a)Infoseek Newswire、Industry、およびPremier（プレミア）の情報源（Infoseekからなる）—Reuters（ロイター）、PR News Wireなど、及び(b)NewsTracker（Exciteからなる）—オンライン新聞・雑誌

4. 画像—画像インデックス：(a)Corel（コレル）—corel画像データベース、(b)HotBot—HotBot画像、(c)Lycos—Lycos画像、(d)WebSeer（ウェブシーア）—WebSeer画像、(e)Yahoo—Yahoo画像、(f)AltaVista—AltaVista画像

5. 雑誌—学術雑誌：(a)Science（サイエンス）

6. テクノロジー—技術ニュース：(a)TechWeb（テックウェブ）及び(b)ZDNet（Z D ネット）

7. 全部—上記全て

制限メニュー16が、検索ボックス14の下にある。制限メニュー16のは、個別のドメイン、個別のページの期間および個別の画像種類に対して結果を制限するためのオプションがある。その次に、結果の最大数、質問用語周囲で表示すべき文脈の量（文字で）、及び、クラスタリングもしくは追跡（トラッキング）を起動させるか否かを選択するための選択肢があるメインオプションメニュー20がある。

40 【0030】一番上のバー14にあるオプションのリンクによって、図2において符号22で示したような多くの他のオプションを設定することができる。そのオプションには、

1. (個々のページのダウンロードごとの) タイムアウト、
2. ビュー（表示）を行う際にページにフィルター処理を行うか否か、
3. ビューを行う際にページからの画像にフィルター処理を行うか否か、
4. 各検索が新規ウィンドウに結果を表示するか否か、

5. (画像のマニュアル操作での分類のため) 画像の分類を行うか否か、

がある。さらに、図2には、質問とURLの変更を追跡し、新たなURLの入力の追跡を可能とするオプションページを、ページ24、26として示してある。

【0031】図3～8には、質問「"nec"および"digital watermark (電子透かし入れ)"」についての、本発明に基づくメタ検索エンジンの応答例を示してある。図3には、検索からの応答の最初の部分を示してある。検索フォームが一番上にあり、その次に、質問によって変わり得るヒント (tip) 30の表示がある。全ての質問用語を含む結果が、検索・解析されるにつれて表示される。前述のように、最初のいくつかのページのいずれもが質問用語の全てを含むものでない場合、検索エンジンは、最初に、それまでのページで認められる最大数の質問用語を含む結果を表示する。ドキュメントタイトルの左にあるバー32は、そのドキュメントで質問用語同士がどの程度近接しているかを示している。すなわち、バーが長いほど、質問用語が互いに近いことを示している。そのドキュメントを検索した検索エンジン、ドキュメントの期間 (ドキュメントがウェブ上に存在してから

の時間)、ドキュメントの大きさ及びURLが、ドキュメントのタイトルに引き続いて表示されている。
【0032】ページの検索を行った後、メタ検索エンジンは、図4に示すように、用語近接度情報を用いて順位付けした最初の20のページを表示する。図5～8で説

* 明するように、下に行くにつれて、メタ検索エンジンは、質問用語が上に表示されているページより少ないページ、質問用語がないページ、重複文脈文字列を含むページ、そしてダウンロードできなかったページを表示する。次に、使用した検索エンジンページへのリンクが表示され、その次に質問の拡張に有用と考えられる用語が表示される。図8について説明すると、メタ検索エンジンは、次に、検索エンジンごとに見いだしたドキュメントの数、検索および処理した数、及び重複数に関する情報がある要約ボックスを表示する。

【0033】図9には、ビュー時に個々のページをどのように処理するか例を示してある。一番上にあるリンク40は、それをクリックすることによってドキュメント内で質問用語が最初に出てくる箇所へジャンプすることができ、また、その用語が出てくる回数を示している。[Track Page (追跡ページ)] リンクは、そのページについての追跡を作動させるものであり、ユーザーは、そのドキュメントについての変更の時期および方法に関する情報を得ることができる。

【0034】このメタ検索エンジンは、2つの主要な論理部分、すなわちメタ検索コードと並列ページ検索デモンを有する。(簡易版の) 検索コードについての擬似コード (pseudocode) は次の通りである。

【0035】

【数2】

```

・要求を処理して構文をチェックし、質問用語と突き合わせるのに使用される正規表現を作り出す。
・全ての関連する検索エンジンに、要求 (適切に変更を加えたもの) を送る。
・結果の最大数または検索された全てのページに達するまで、検索した各ページについてループ (#ループ終わりまで) を行う。
{
  ・ページが検索エンジンからのものである場合、
  {
    ・検索エンジンの応答を解析して、ヒットおよび結果に関する次の集合へのついでリンクを放散する。
    ・全てのヒットについて要求を送る。
    ・可能であれば、次の結果集合について要求を送る。
  }
  ・ページが検索エンジンからのものでない場合、
  {
    ・質問用語についてページをチェックし、見つければ文脈文字列を作る。
    ・全ての質問用語が見いだされ、重複文脈文字列が以前になかった場合、ページ情報と文脈文字列を印刷する。
  }
}
#ループ終了
終了ループ
・近接度および用語頻度情報を用いてページの再順位付けを行う。
・一部の質問用語を含むが全ては含まなかったページについて、ページの情報及び文脈文字列を印刷する。
・質問用語を含まなかったページについて、ページの情報および文脈文字列を印刷する。
・重複文脈文字列を含むページについて、ページの情報および文脈文字列を印刷する。
・ダウンロードできなかったページについて、ページの情報および文脈文字列を印刷する。
・統計解析の要約を印刷する。

```

図10は、メタ検索エンジンの制御フローチャート50を簡略化して示している。このページ検索エンジンは、

比較的に簡潔なものであるが、要求の待ち合わせ (キューイング) 及び複数検索プロセスからの負荷の平準化、な

らびにサイトへの過負荷を防止するための同一サイトへの要求の遅延などの特徴を組み込んでいる。以下、このフローチャートを説明する。

【0036】メタ検索エンジンは、インターネット上の各ウェブ（サイト）202を並列して検索する並列ページ検索エンジン201を含んでいる。ウェブ202には、サードパーティの検索エンジンも含まれている。図10において破線の矢印は、データの流れを示している。

【0037】ユーザが質問(query)を入力すると（ステップ211）、各検索エンジンに併せて質問を手直しし、その質問を各検索エンジンに送信する（ステップ212）。その後、予め定めた最大ヒット数未満であってかつ検索するページが残存しているかを判断し（ステップ213）、そうでない場合にはステップ223に移行し、予め定めた最大ヒット数未満であってかつ検索するページが残存している場合には、各検索エンジンでのページの検索を待ち合わせ（ステップ214）、検索エンジンからページが入力したかどうかを判断する（ステップ215）。

【0038】ステップ215でページが入力している場合には、その検索エンジンからの応答を解析し（ステップ216）、その解析結果に基づいて、ウェブページの要求を送信し（ステップ217）、検索エンジンでさらにヒットしたかを判断する（ステップ218）。ここでヒットしなかった場合には、次のページの検索のために＊

```

・クライアント始動
・ループ開始（#ループ終了まで）
{
  ・アクティブクライアントのタイムアウトをチェックする。
  ・可能であれば、待ち合わせした要求を送り、複数の検索プロセスからの要求について負荷の平均化を行う。
  ・クライアントからメッセージがある場合、
  {
    ・メッセージが「切り換え（replace me）」である場合、クライアントを新たなプロセスに切り換える。
    ・メッセージが「完了（done）」である場合、クライアント情報を更新する。
    ・メッセージが「ステータス（status）」である場合、ステータスを返す。
    ・メッセージが「取得（get）」である場合、全てのクライアントがビジー状態であるか、あるいは最近x秒以内にそのサイトに対して要求を行っていた場合には要求を待ち合わせし、そうでなければ、クライアントに要求を送る。
  }
}
#ループ終了

```

クライアントプロセスは、単に、関連するページを検索して、エラーとタイムアウトを扱い、適切な検索プロセスに、直接、そのページを送り返すものである。

【0044】さらにこのメタ検索エンジンでは、質問用語に関するメタ情報のデータベースを作成して、そのデータベースを使用することにより、ユーザにとってより適切な情報をユーザに提示することができる。例えば、映画のタイトル（題名）についてのリストをメタ検索エンジンが保持するようにしておき、ユーザが映画のタイトルを含む質問を入力したときには、メタ検索エンジン

＊ステップ213に戻り、ヒットした場合には、ヒットした結果に含まれる、次のページの要求を送信して（ステップ219）、ステップ213に戻る。

【0039】ステップ215でページが入力していない場合には、既に検索したドキュメントを分析し（ステップ220）、そのドキュメントが表示の基準を満たしているかを判断して（ステップ221）、満たしていない場合には、次のページの検索のためにステップ213に戻り、満たしている場合には、質問用語文脈とともにドキュメントを表示してから（ステップ222）、ステップ213に戻る。

【0040】ステップ213で、最大ヒット数に達したか、検索するページが残っていない場合には、結果を表示する段階に入り、異なる順位付け基準とともに、結果を表示し（ステップ223）、前の表示基準には適合しなかった結果を表示し（ステップ224）、統計解析の要約を表示して（ステップ225）、処理を終了する。

【0041】以上述べた、メタ検索エンジンは、実際には、コンピュータ上にソフトウェアとして実装されるものであり、特に、図10に示した制御処理は、コンピュータ上で実行されるプロセスの一種であるデーモンとして実行されるものである。

【0042】（簡易版の）ディスパッチデーモンについての擬似コードは次の通りである。

【0043】

【数3】

40 がその質問を認識してリストを参照し、インターネット上の特定の映画評論サイトにある当該映画についての評論のページにアクセスして、そのページをユーザに提示するようにすることができる。

【0045】本発明のメタ検索エンジンにおいて画像のメタ検索に使用されるアルゴリズムは、以下の通りである。擬似コードで示す。

【0046】

【数4】

15

```

・要求を処理して構文をチェックし、質問用語と突き合わせるのに使用される正規表現
  を作り出す。
・全ての関連する検索エンジンに、要求（適切に変更を加えたもの）を送る。
・画像の最大数または検索された全てのページに達するまで、検索した各ページについ
  てループ（#ループ終わりまで）を行う。
{
  ・ページが検索エンジンからのものである場合、
  {
    ・検索エンジンの応答を解析して、ヒットおよび結果に関する次の集合へのついで
      リンクを放料する。
    ・全てのヒットについて要求を送る。
    ・可能であれば、次の結果集合について要求を送る。
  }
  ・ページが検索エンジンからのものでないが画像である場合、
  {
    ・画像を表示キューに加える。
  }
  ・ページが検索エンジンからのものでなく、かつ画像でもない場合、
  {
    ・ページにおける質問用語の位置を解析し、ページにあるどの画像が（該当する
      ものがあれば）その質問に対応するかを予測し、その画像をダウンロードするた
      めの要求を送る。
  }
  ・n個の画像が表示キューにある場合、
  {
    ・キューにある画像の1個の合成画（モンタージュ）を作る。
    ・オリジナルの個々の画像に相当する画像の各部分がその元の画像についての詳
      細ページを示すクリック可能な（クリックابل）画像として、合成画を表示す
      る。
  }
}
#ループ終了
・画像が表示キューにある場合、
{
  ・その持ち合わせにある画像の1個の合成画を作る。
  ・オリジナルの個々の画像に相当する画像の各部分がその元の画像についての詳細
    ページを示すクリック可能な（クリックابل）画像として、合成画を表示する。
}
・統計解析の要約を印刷する。

```

16

図11は、画像メタ検索エンジンアルゴリズムについての制御フローチャート60を簡略化して示している。以下、このフローチャートを説明する。

【0047】画像メタ検索エンジンは、インターネット上の各ウェブ（サイト）202を並列して検索する並列ページ検索エンジン201を含んでいる。ウェブ202には、サードパーティの検索エンジンも含まれている。図11において破線の矢印は、データの流れを示している。

【0048】ユーザが質問(query)を入力すると（ステップ231）、各検索エンジンに併せて質問を手直しし、その質問を各検索エンジンに送信する（ステップ232）。その後、予め定めた最大ヒット数未満であってかつ検索するページが残存しているかを判断し（ステップ233）、そうでない場合にはステップ247に移行し、予め定めた最大ヒット数未満であってかつ検索するページが残存している場合には、各検索エンジンでのページの検索を待ち合わせ（ステップ234）、検索エンジンからページが入力したかどうかを判断する（ステップ235）。

【0049】ステップ235でページが入力している場合には、その検索エンジンからの応答を解析し（ステッ

30 プ236）、その解析結果に基づいて、ウェブページの要求を送信し（ステップ237）、検索エンジンでさらにヒットしたかを判断する（ステップ238）。ここでヒットしなかった場合には、次のページの検索のためにステップ233に戻り、ヒットした場合には、ヒットした結果に含まれる、次のページの要求を送信して（ステップ239）、ステップ233に戻る。

【0050】ステップ235でページが入力していない場合には、既に受け取ったページが画像かどうかを判断し（ステップ240）、画像でない場合には、質問用語でそのページを検索し、画像に合致すると考えられる質問を予測し（ステップ241）、質問に合致すると予測される画像の要求を送信して（ステップ241）、ステップ213に戻る。また、ステップ240でページが画像である場合には、そのページが表示の基準を満たしているかを判断して（ステップ243）、満たしていない場合には、次のページの検索のためにステップ233に戻り、満たしている場合には、その画像を表示キューに追加してから（ステップ244）、表示キューが一杯かどうかを判別し（ステップ245）、表示キューが一杯でなければそのままステップ233に戻り、表示キューが一杯であれば、画像のモンタージュを作成・表示し、

表示キューをクリアしてから(ステップ246)、ステップ233に戻る。

【0051】ステップ233で、最大ヒット数に達したか、検索するページが残っていない場合には、結果を表示する段階に入り、表示キュー中の画像のモンタージュを作成・表示し、統計解析の要約を表示して(ステップ247)、処理を終了する。

【0052】以上述べた、画像メタ検索エンジンは、実際には、コンピュータ上にソフトウェアとして実装されるものである。

【0053】画像の分類: ウェブ画像検索エンジンWebSeerは、写真またはグラフィックとして画像の分類を試みるものである。WebSeerは、画像から多くの特徴を抽出し、分類のために決定木を使用する。本発明者らは、同様の画像分類システムを実装した。しかしながら本発明者らは、異なる特徴集合を用い、ニューラルネットワークを使って分類を行う。図12及び図13は、写真についてフィルター処理を行った画像を用いての、画像質問「コアラ(koala)」に対する本発明のメタ検索エンジンの応答を示している。図14は、グラフィックにつ

いてフィルター処理を行った場合の応答を示している。
【0054】ここで分類項目としては、例えば、写真、線画、ロゴ、地図、漫画、ポーレート、ボタン、チャート及び天文画などが挙げられる。フィルター処理を行う場合には、画像の種類が、ここで挙げた分類項目のうちのいずれに該当するかをまず同定し、その同定結果に基づいて、フィルター処理を実行する。その際、画像の種類などのメタ情報を保存することにより、新たな質問が入力したときに、画像の再ダウンロードを実行することなく、メタ情報を用いて画像のフィルター処理を実行

することができるようになる。

【0055】ドキュメントのクラスタリング: ドキュメ

- ・質問に相当するページを検索する。
- ・各ページについて、
 - {
 - ・ $n=1$ から最大フレーズ長さ(MaximumPhraseLength)までについて、
 - {
 - ・ 逆読する n 個の単語からなる各集合について、この単語組み合わせが、このドキュメントでまだ出ていなかった場合、その集合を、その文書についてのハッシュテーブルと全ドキュメントについてのハッシュテーブルに加える。
 - }
 - }
- ・ n = 最大フレーズ長さから1までについて、
 - {
 - ・ Min回より多く出てきたフレーズであって、最大、Max個のフレーズまで、長さ n の最も一般的なフレーズを見いだす。
 - ・ それらのフレーズをクラスターの集合に加える。
 - }
- ・ 以前のステップから、最大Max個の組み合わせまで、個々のドキュメント中で組み合わせとしてMin回以上出てきた最も一般的な2個のクラスターの組み合わせを見いだす。
- ・ 別のクラスターを同定するフレーズの部分集合であるフレーズによって同定されるクラスターを削除する。
- ・ 同じドキュメントを含むクラスターをマージする。
- ・ 質問用語とクラスター用語の両方についてのページ集合からの文脈とともに、各クラスターを表示する。

メントのクラスタリング法は、典型的には、重複しないクラスターを生み出すものである。例えば、ドキュメントのクラスタリングで最も一般的に使用されるアルゴリズムである階層集塊クラスタリング(Hierarchical Agglomerative Clustering; HAC) (Willett, P., "Recent trends in hierarchical document clustering: a critical review", Information Processing and Management, 24, 577-597(1988)) は、あるクラスター中の各ドキュメントから始めて、停止条件が満足されるまで、クラスターを繰り返しマージする。HACアルゴリズムは、文書間及び文書集合間の類似度関数を使用している。

【0056】本明細書は、同時に出てくるフレーズ(phrase)およびフレーズの結合の同定に基づいた文書クラスタリングアルゴリズムが開示している。そのアルゴリズムは、一般に使用される方法とは、クラスターの重複があり得て、共通の項目もしくはテーマを同定するためのものである点で、基本的に異なっている。

【0057】WWW(World Wide Web; ウェブ)は大きく、多くの冗長性を有し、シグナル対ノイズ比(S/N)が比較的小さい。これらのファクターは、ウェブ上での情報検出を困難にしている。本明細書で提供するクラスタリングアルゴリズムは、情報発見を支援するためのものである。すなわち、ある質問について返ってきた多くのヒットから、どのトピックが含まれているかを発見するためのものである。これによってユーザーは、質問に改善を加えて、小トピックの一つを調べることができる。

【0058】そのクラスタリングアルゴリズムは次の通りである。擬似コードで示す。

【0059】

【数5】

る。図16は、これらクラスターについての最初の2つのクラスターの要約72、74を示している。図17及び図18は、同じ質問についてHuskySearch（ハスキーサーチ）によって形成されたクラスター76、80を示している。図19は、AltaVistaによって形成されたクラスター82を示している。図20及び図21は、別の2つの質問「ニューラルネットワーク（neural network）」および「typing and injury」について、本発明のメタ検索エンジンによって形成されたクラスター84、86を示している。

【0060】質問拡張：質問の拡張を行う一つの方法は、質問用語の形態を変えることで質問を大きくするのである。単語語幹化（stemming）（Porter, M.F., "an algorithm for suffix stripping", Program, 14, 130-137(1980)）を用いて、単語の変形語を同一単語として処理することができる。ウェブ検索エンジンは、ウェブのインデックス付けを行うのに必要なリソースを減らすと考えられているにもかかわらず、単語語幹化を行わないのが普通である。ウェブ検索エンジンが単語語幹化を行わない理由の一つは、語幹化によって精度が低下する

可能性があるというものである。語幹化は、全ての変形語を考慮するものである。全ての変形語を用いる質問拡張では、変形語が異なる概念を指す場合が多いことから、ウェブ検索の精度を低下させることになる場合が多い。ウェブのデータベースがより大きく、より多様になっていることから、単語語幹化を用いた場合の精度低下は、従来の情報検索テストを集成した場合と比較して、ウェブでは問題となりやすいのが普通である。

【0061】本明細書では、変形語の部分集合のみの使用に基づいた質問拡張アルゴリズムが開示されている。具体的には、そのアルゴリズムでは、元の質問に合致するウェブページのうちの一定のパーセントのウェブページにおいて出てくる変形語の部分集合を用いる。現在、質問用語は、ポーターのステマー（stemmer）（Porter, M.F., "An algorithm for suffix stripping", Program, 14, 130-137 (1980)）によって語幹化されており、質問用語の変形語について、検索ページを検索することができる。それらのページの1%より多くで出てくる変形語をユーザーに表示して、それ以降の質問に含めることができる。この方法についての定量的な評価は行っていないが、有用な用語が提案されるのが認められる。例を挙げれば、「nec」と「digital watermark」という質問については、digitally, watermarking, watermarks, watermarkedという用語が、質問拡張のためのものとして提案される。

【0062】質問拡張用語は質問が完了するまで未知であることから、現在、この技術は、最初に質問を入力した時に、質問を自動的に拡張することはない。しかしながら、各質問用語について拡張用語のデータベースを維持することで、この技術を自動化することができる。あ

る用語を含む最初の質問によって、同時に出てくる変形語をデータベースに追加することができ、それ以降の質問ではそれらの用語を使って必要に応じてデータベースを更新することができる。

【0063】具体的な表現形式：情報が多くの形で表わされる可能性があって、意味論（semantic）の組み込みおよび自然言語の理解を行う最適な検索システムを必要とするため、正確な情報検索は困難である。情報検索における研究では、例えば単語語幹化および質問拡張などの再検索能力向上を目的とした技術を考慮する場合が多い。前述のように、これらの技術によって、特にウェブと同じように多様なデータベースでは、精度が低下する可能性がある。

【0064】WWW（World Wide Web）は多くの冗長性を有する。情報は複数回含まれることが多く、ウェブを通じ、各種形式で表現される。全ての情報が全ての可能な形で表現されるという制限下では、高精度の情報検索は簡単なものとなり、意味論解析上の知識を必要としないと考えられる。そこで、ある特定の情報表現方法について検索する必要があるだけとなると考えられる。そのような目標は、全ての情報については決して達成されるものではないが、実験から、ウェブはすでに、この考え方に基づいたアプローチを行えるものとなっており、ある種の検索作業には有効であることが示されている。

【0065】本発明の方法は、疑問の形の質問を、回答を表現するための具体的な形式に変換するものである。例えば、「NASDAQは何を表すか（What does NASDAQ stand for?）」という質問は、「NASDAQは表わす」、「NASDAQは略称である」、「NASDAQは意味する」という質問に変換される。明らかに、その情報は、これら3つの可能性に対して異なった形で含まれると考えられる。しかしながら、その情報がこれら形式のいずれかに存在する場合、それらのフレーズを見出すことで、質問に対する回答が得られる確率が高くなる。この手法は、再検索によって精度を売るものである。

【0066】本発明のメタ検索エンジンは、現在、以下の質問について、固有表現形式（SEF: specific expression forms）手法を用いている（大括弧[]は選択肢を示し、括弧()はオプションの用語または選択肢を示す）。また、{}内は、対応する英文での質問を示している。

【0067】・x[単数|複数]とは何か。{What [is|are] x?}

・xは何によって[生じるか|作られるか|形成されるか]。{What [causes|creates|produces] x?}

・x[について|のことを|に関して]あなたは思うか。{What do you think [about|of|regarding] x?}

・xは何を[表すか|意味するか]。{What does x [stand for|means]?}

・xはどこにあるか。{Where is x?}

- ・ xは誰か。{Who is x?}
- ・ (| その) x [単数 | 複数] は [なぜ | どのようにして] yであるか。{[Why|how] [is|are] (a|the) x y?}
- ・ なぜxを行うか。{Why do x?}
- ・ xはいつか。{When is x?}
- ・ いつxを行うか。{When do x?}
- ・ どのように、x [を行うか | 行うことができるか]。{How [do|can] I x?}
- ・ [| その] xはどのようにしてyを行う(ことができる)か。{How (can)[a|the x y?]}
- ・ [| その] xはどのようにしてyを行うか。{How does [a|the] x y?}

【0068】変換の例として、「xは何を[表すか | 意味するか]」は、「xは表す{x stands for}」、「xは略称である{x is an abbreviation}」、「xは意味する{x means}」に変換され、「xは何によって[生じるか | 作られるか | 形成されるか]」は、「xを生じる{x is caused}」、「xは作られる{x is created}」、「xをもたらす{cause x}」、「xを形成する{create x}」、「xを作る{make x}」に変換される。

【0069】異なる検索エンジンは、異なった停止語(stop word)および適合度尺度を使用し、それにより、一部の検索エンジンは、SEFを含まない多くのページを検索結果として与える傾向がある。したがって、関連するエンジンについて、望ましくない文面が質問から除去される。

【0070】図22は、符号90に、「NASDAQは何を表すか」という質問についての本発明のメタ検索エンジンの応答を示している。この質問に対する回答は、最初の6個のページからの約5個のページについて表示されたローカル文脈に含まれている。図23は、符号92に、同じ質問に対するInfoseekの応答を示してある。この質問に対する回答はページ要約には表示されず、回答があった場合にどのページが回答を含んでいるかははっきりしない。図24及び25は、それぞれ、符号94と符号96に、「虹はどのようにして作られるか(How is a rainbow created?)」という質問に対する本発明のメタ検索エンジンとInfoseekの応答を示してある。やはりその回答は、本発明のメタ検索エンジンによって示されたローカル文脈に含まれているが、Infoseekでは、それが挙げたページのどれが質問に対する回答を含んでいるかははっきりしない。図26は、符号100に、「ミーリー機械とは何か(What is a Mealy machine)」という質問についての本発明のメタ検索エンジンからの応答の第3の例を示している。

【0071】当然のことながら、容易にアクセスできる情報の量は、時間の経過に伴って増加することから、固有表現形式手法(SEF)の表現形式方法の実行可能性(viability)は、時間の経過に伴って向上すると予想

される。上記で議論の手順の延長として、各種SEFについて順序が決定される。例えば「xは表す」は、「xは意味する」というフレーズより「xは何を表すか」というフレーズに対して回答を見出しやすいと考えられる。いずれのSEFも見出し出されない場合は、メタ検索エンジンは、標準的質問に戻る。

- ・ 【0072】このメタ検索エンジンによって、検索ヒント(search tips)を得ることができる。そのヒントとして表示される文面には例えば、次のものが含まれるものと考えられる。

【0073】・例えば、「nec research」というように、フレーズについての引用符を用いるとよいでしょう。

【0074】・あなたは、「隠す(hide)」リンク上をクリックすることで、画面スペースを節約するために上記の種々の選択肢を隠すことができます。

【0075】・ウィンドウオプション：ヒットをクリックすることで、多重検索のために同一ウィンドウに、あるいは各新規検索のために新たなウィンドウに、ページが表示されます。

- 20 【0076】・フィルターオプション：ビュー時にページにフィルター処理を行って、質問用語を強調します。そのページのローカルキャッシュ化により、表示が高速化します。

【0077】・ページタイトル後の文字は、結果を与えた検索エンジンを識別するものです(例：A=AltaVista)。

【0078】・ページタイトル後の2番めのフィールドは、そのページが最後に更新されて以降の時間を示します(例：5m=5ヶ月、1y=1年間)。

- 30 【0079】・ページタイトル後の3番めのフィールドは、ページのサイズです。

【0080】・文脈オプションは、質問用語の前後いずれかの側を表示する際の文字数を選択するものです。

【0081】・タイムアウトオプションは、各個々のページをダウンロードする際の最大時間を設定するものです。

【0082】・「Press(報道機関)」での検索は、現在のニューストピックでより高い精度を得る上で有用です。

- 40 【0083】・画像オプション：より迅速にビューを行うため、ビュー時にページから画像を除去します。

【0084】・フィルター処理されたページのビューを行う場合、質問用語をクリックすることで、その用語が次に出てくる所にジャンプすることができます。最後に出てきた用語をクリックすることで、最初に出てきた所にジャンプして戻ります。

【0085】・「-term」オプションを用いて、用語を除外することができます。

- 50 【0086】・特定ページへのリンク、例えばリンク:ww.neci.nj.nec.com/homepages/qilesiについて検索を行

うことができます。ただし、自己リンクは除外されます。

【0087】・疑わしい場合は、小文字を使用した方がよいでしょう。

【0088】・このメタ検索エンジンは、単独の検索エンジンの場合の3倍を超えるドキュメントを利用できるようにするものです。例えば、NASDAQが何を表すかを知らたい場合、検索に限定を加えることで、「NASDAQ」より「NASDAQは表す」について検索を行う方が、より迅速に回答を見いだすことができます。ただし、その情報も別の形で表現されている場合もあります。

【0089】・「検索：(Searching for:)」の行にある検索エンジンリンクをクリックすることで、現在の質問に対するその検索エンジンの応答が示されます。

【0090】・「画像(images)」ボタンを選択することで、画像、例えば「red rose」の検索を行うことができます。

【0091】・タイトルの左にあるバーは、質問用語がドキュメント中で互いに近くなるほど長くなります。

【0092】・「検索」行における質問用語リンクをクリックすると、Webster (ウェブスター) 辞書でのその用語の定義を表示します。

【0093】・「追跡：行う(Tracking: Yes)」を選択すると、質問が追跡され、「NEC検索についての最近の項目(recent articles about NEC Research)」と同様のあなたのためにカスタマイズされたホームページに、新たなヒットが表示されます。

【0094】・「クラスタリング：行う(Cluster: Yes)」を選択して、ドキュメントのクラスタリングを行い、共通のテーマを確認することができます。

【0095】・「画像：(Images:)」オプションを用いることで、各画像が写真であるかグラフィックであるかのニューラルネットワーク予測を行い、画像にフィルターをかけることができます。

【0096】・全てのドキュメントの検索が終了した後、用語の近接度によって順位付けしたページのリストが表示されます。

【0097】以上が、このメタ検索エンジンにおいて、ヒントとして表示される文面の例である。さらには、ユーザが入力した質問に基づいて文脈に応じた提案を行うようにすることもできる。例えば、入力された質問がイニシャルを表わすと考えられる単一の文字を含む場合に、名前についての検索を行う方法についての提案あるいは示唆を表示するようにすることもできる。

【0098】質問およびURLの追跡

Informant (The Informant, 1997) (インフォマント) などのサービスは、質問に対するウェブ検索エンジンの応答を追跡し、新たなドキュメントが発見されたらユーザーに知らせるものである。本発明のメタ検索エンジンは、この機能をサポートする。質問実行時に追跡(Trac

k) オプションを選択することで、その質問についての追跡が起動される。次に、デーモンが、質問を定期的に繰り返して、新たなドキュメントをそれが発見された時刻とともに保存する。新たなドキュメントは、図27の符号102に示したように、メタ検索エンジンのホームページ上でユーザーに呈示される。質問に一致するドキュメントが変更されているか否かについては、このメタ検索エンジンは現在のところ、ユーザーに知らせない。ただし、それを追加することは可能であると考えられる。

【0099】本発明のメタ検索エンジンはさらに、URLの追跡もサポートする。検索エンジン結果からのページのいずれかのビューを行う際に、[ページの追跡(Trackpage)] リンクをクリックすることで、追跡を開始する。別法として、オプションページを用いて、任意のURLについての追跡を起動することができる。デーモンは、追跡対象ページに対する更新を確認し、図27に示すように、ホームページ上で、ユーザーに対して変更のあったページのリストを示す。[ページ(Page)] リンクは、図28に例示するように、追跡対象ページを表示し、最初の部分にヘッダを挿入して、ユーザーがそのページを最後に見て以降、どの行に追加もしくは変更があったかを示す。

【0100】検索エンジンの範囲とウェブのサイズの推定：WWW (World Wide Web) は拡大し続けていることから、科学者にとって徐々に重要な情報源となりつつある。全ての科学文献を即座に閲覧することは、科学者の長年の夢であり、ウェブ検索エンジンは、科学文献および他の情報源の大きくかつ成長し続けている閲覧容易な集合体を形成するようになった。主要なウェブ検索エンジンは、一般に、ウェブのかなりの割合の部分に対してインデックス付けを行うと考えられている。検索方法の選択に影響を与える重要な疑問には、「検索エンジンは、ウェブのどの部分についてインデックス付けを行うか」、「どの検索エンジンが最も包括的であるか」、「検索エンジンのデータベースは、どの程度最新のものか」などがある。

【0101】検索エンジンについての多くの比較が利用できる。それには、多くの検索エンジンでの1組の質問実行と、各検索エンジンが返してきた結果数の報告とが含まれるのが普通である。検索エンジンは、質問用語の入っていないドキュメントを返す場合があることから、それらの比較の結果の信頼性は限られている。それは、(a)エンジンが使用する情報検索法(例えば、Exciteは「概念に基づくクラスタリング」を使用し、Infoseekは形態論を用いる。これらのエンジンは、関連する単語を含むドキュメントを返してくる場合がある)、(b)ドキュメントがもはや存在しない場合があること(無効なドキュメントを決して削除しないエンジンが有利になると考えられる)、(c)ドキュメントは存在するものの、変

更されていて、質問用語をもちや含まない場合があること、が原因となるものと考えられる。

【0102】セルバーグ及びエチオニ (Selberg, E. and Etzioni, O., "Multi-service search and comparison using the MetaCrawler", Proceedings of the 1995 World Wide Web Conference, (1995)) は、MetaCrawler メタ検索サービスの利用ログに基づいた結果を発表している (検索エンジンサービスとウェブにおいてかなりの変更があったため、その結果は、現在、同じことを繰り返したとしたらかなり異なるものになると予想される)。これらの結果は、Lycos、WebCrawler、InfoSeek、Galaxy (ギャラクシー)、Open Text (オープンテキスト) 及びYahooという検索エンジンを検討したものである。セルバーグ及びエチオニの結果は、有益ではあるが、いくつかの理由により制限がある。

【0103】第1に、セルバーグらは、各検索エンジンの「市場シェア (market share)」を提供しており、それは、各検索エンジンから得られたドキュメントでユーザーがフォローするものパーセント (割合) である。(a) ページを見なければ関連性を決定することは困難であること、(b) ドキュメントを提供する順序がユーザーの関連性判断に影響を与えること (Eisenberg, M. and Barry, C., "Order effects: A preliminary study of the possible influence of presentation order on user judgments of document relevance", Proceedings of the 49th Annual Meeting of the American Society for Information Science, Vol. 23, pp. 80-86 (1986))、等の多くの理由のために、セルバーグらの結果は、限定的なものとなる。

【0104】セルバーグ及びエチオニによって検討された結果は、さらに、彼らが、特定の1つの検索エンジンのみから検索されたドキュメントのパーセントと、各検索エンジンの検索範囲とによって、結果を表わしていることから限定的なものである。彼らの結果は、各検索エンジンがウェブの一部のみカバーすることを示唆するものである。しかしながら彼らの結果は、(a) 上述のように、検索エンジンは質問用語が含まれていないドキュメントを返す場合があり、関連する単語があるドキュメントまたは無関係なドキュメントを返してくる検索エンジンによって、かなり異なる結果が生じる場合があること、(b) 検索エンジンは異なった順序でドキュメントを返してくるため、妥当な比較を行うには全てのドキュメントを検索する必要があること、例えば、2つの検索エンジンがドキュメントの全く同一の組み合わせにインデックス付けを行うが、最初のxとして異なる組み合わせを返してくることがあること、等の理由により、限定的なものである。

【0105】さらに、セルバーグおよびエチオニは、無関係なリンクのパーセントが15%であることを見出している。彼らはそれを、検索エンジンごとに分析してい

ない。セルバーグおよびエチオニは、その研究における限界を指摘している (非常に好結果を示したMetaCrawler サービスについての膨大な論文のごくわずかな部分を割いて)。

【0106】検索サービスAltaVistaおよびInfoseekは、最近、それらがウェブを総合的に網羅していないことを確認している (Brake, D., "Lost in cyberspace", New Scientist, 154(2088), 12-13, (1997))。以下に、これら検索エンジンがどの程度網羅しているかの推定について、議論する。

【0107】本発明者らは、主要なウェブ検索エンジンの検索範囲、ウェブのサイズ、検索エンジンのデータベースの最新性についての統計解析を行った。ここでは、AltaVista、Excite、HotBot、Infoseek、LycosおよびNorthern Light (アルファベット順) という6つの最近の主要なフルテキスト (全文) 検索エンジンのみについて検討した。一般的認識は、これらの検索エンジンが、同一ドキュメントを大まかにインデックス付けし、ウェブの比較的大きい部分についてインデックス付けを行うというものである。

【0108】本発明者らは最初に、1~6個の検索エンジンの異なる組み合わせを用いた場合に、検索結果として返されてくるドキュメントの数を比較する。全体を通じての本発明者らの方法は、全ての検索エンジンからの合致ドキュメントのリストを検索し、次にそれらドキュメント全てについて検索を行って解析するというものである。2つの重要な制約を用いた。

【0109】第1の制約は、ある質問を試験に含めるには、その質問に合致するドキュメントの全体リストを、全ての検索エンジンについて検索していなければならないというものであった。検索エンジンがドキュメントに順位付けを行う順序が、検索エンジン間で異なることから、この制約は重要である。各検索エンジンから1000を超えるドキュメントが得られた質問について検討する。各検索エンジンからの最初の200のドキュメントのみを比較していたら、特定の1つの検索エンジンからの結果にのみ見られるURLを多く見いだすことができる。しかしながら本発明者らは、それら検索エンジンが、固有のURLのインデックス付けを行ったか、あるいは、それら検索エンジンが同一のURLのインデックス付けを行ったが最初の200個のドキュメントとして異なる部分集合を返したか、について決定することができないと考えている。

【0110】第2の制約は、本発明者らが、各検索エンジンが質問に合致するとしてリストアップするドキュメント全てについて、対応するURLの全テキストをダウンロードすることを試みたというものである。ダウンロードできて、実際に質問用語を含むドキュメントのみを数える。これは、(a) 検索エンジンの中には、その検索エンジンは関連があると考えるが、実際には質問用語を

含んでいないドキュメントを返してくるものがあること（例えば、Exciteは「概念に基づくクラスタリング」を用いるとともに関連する単語を考慮する場合があります、Infoseekは形態論を用いる）、(b)各検索エンジンは多くの無効なリンクを含み、無効なリンクのパーセントは検索エンジン間で異なること（無効なリンクを削除しない検索エンジンが有利になると考えられる）から、重要である。

【0111】解析において重要な他の詳細事項としては、以下のものがある。

【0112】1. 異なるURLを有する同一ページの検出など、1個の検索エンジンまたは複数の検索エンジンの組み合わせによって返されてくるドキュメントの総数を検討することで、重複が除かれる。URLは、a)その末尾に配置される「index.html」接尾語または終端文字「/」を除き（一般には、URLとしてhttp://www.aaa.comとhttp://www.aaa.com/とhttp://www.aaa.com/index.htmlを入力すると、同じウェブページが表示される）、b)ポート番号80の指定（デフォルト）を除き（TCP/IP(Transmission Control Protocol/Internet Protocol)において、ポート番号80は、ウェブページ表示用のプロトコルであるhttp(Hyper Text Transport Protocol)を指す）、c)1より大きいディレクトリ深度(directorydepth)を有するURLのドメイン名の最初のセグメントを除き（マシンのエイリアスを考慮するため）、d)エスケープ文字（例：URL中の"%E"は、波形文字("-")と同等である）による拡張を除くことで、規格化される。

【0113】2. 検索エンジンが異なると、大文字による質問を異なった形で処理することから（例：AltaVistaは、大文字の質問に対しては大文字の結果のみ返してくる）、本発明者らは、小文字の質問のみを考慮する。

【0114】3. 本発明者らは、個々のページのタイムアウトとして60秒を用いた。タイムアウトとなったページは解析には含まなかった。

【0115】4. 本発明者らは、質問当たりの最大ドキュメント数（重複を除いた後に合わせた全ての検索エンジンからのもの）を700に固定して、それより多いドキュメントを返す質問を含まないようにした。これら検索エンジンは、検索可能な最大数のドキュメントを課すのが普通であり（現在の限界は、AltaVistaで200、Infoseekで500、HotBotで1000、Exciteで1000、Lycosで1000、Northern Lightで10000超である）、本発明者らは、これらの範囲を超えていないことを調べた（この制約を用いると、各検索エンジンから最大数を返して返ってくる質問はなかった。特に、AltaVistaから200を超えて返ってくる質問はなかった）。

【0116】5. 本発明者らは、正確な質問用語を含むドキュメントのみを数えた。すなわち、あるドキュメント中の「crystals」という単語は「crystal」という質問用語とは合致せず、そのドキュメントを質問に合致するものとして数えるには、その単語の単数形がドキュメント中になければならないことになる。これは、検索エンジンが異なると形態規則が異なることから必要である。

【0117】6. HotBotとAltaVistaは、ウェブ上の同一情報を有する別のページを確認する場合がある。これらの別ページは統計解析に含める（それらは同一データを有する別ページを確認しない検索エンジンの場合と同様である）。

【0118】7. Northern Lightの「特殊収集(special collection)」（公的にインデックス付け可能なウェブの一部ではないプレミアのドキュメント）は使用しなかった。

【0119】ある期間にわたって本発明者らは、上記制約を満足する500の質問を用いて、各検索エンジンからの結果を収集した。以下に述べる結果は、本発明者らが、1997年8月23日～1997年8月24日の期間で、500の質問を行って得たものである。検索エンジンは、ドキュメントのリストアップ及び/または次ページのドキュメントの要求についてのフォーマット（書式）を定期的に変更することから、テストの前後で各検索エンジンからの全ての結果が正しく検索および解析されていることを、本発明者らは、マニュアル操作で調べた（本発明者らは、さらに、検索エンジンの応答フォーマットにおける一時的な不首尾や変化を検出するための自動的方法も用いる）。

【0120】図29は、6個の検索エンジンから検索されたドキュメントの全体に対して、各検索エンジンのそれぞれから検索されたドキュメントの総数がどの程度の割合になるかを示している。以下表1に、95%信頼区間とともにこれらの結果を示す。この比較では、HotBotが最も網羅的である。これらの結果は、実行した特定の質問、ならびにそれを行った時点での検索エンジンのデータベースの状態に固有のものである。さらにその結果は部分的には、データベースサイズが異なるためではなく、インデックス付けが異なるためである可能性がある。すなわち、検索エンジンが異なると、同じドキュメントに対して同一の単語をインデックス付けしない場合がある（例えば、これらエンジンは最大ファイルサイズを課して、大きすぎるドキュメントを効果的に切り取るのが普通である）。

【0121】

【表1】

検索エンジン	HotBot	Excite	Northern Light	AltaVista	Infoseek	Lycos
6 個のエンジン についての範囲	39.2%	31.1%	30.4%	29.2%	17.9%	12.2%
95%信頼区間	±1.4%	±1.2%	±1.3%	±1.2%	±1.1%	±1.1%

図30は、6個全ての検索エンジンから検索された全ドキュメント数によって規格化した、1個乃至6個の検索エンジンによって検索されたドキュメント数の平均割合を示してある。1個乃至5個の検索エンジンを用いた場合については、平均は、検索エンジンのあらゆる組み合わせ全体にわたるものであり、各質問について平均を取ってから、質問全体にわたって平均を取ったものである。

【0122】検索エンジンの数とともに検索範囲は対数的に増えるという仮定、ならびにその制限の下では、無限数の検索エンジンによってウェブ全体を網羅することになるという仮定を用いると、 $f(x) = b(1 - 1/\exp(ax))$ (式中、 a および b は定数であり、 x は検索エンジン数である) がデータに適合することになる (プログラム gnuplot におけるデフォルトパラメータを用いた Levenberg-Marquardt 最小化 (Fletcher, R., "Practical Methods of Optimization", Second Edition, John Wiley & Sons, (1987)) を実施)。その結果を図30に示してある。これは、各検索エンジンがウェブのある一定のパーセントを網羅し、各検索エンジンから得られるウェブのサンプルが、全てのウェブページから独立に抜き出される ($c_i = c_{i-1} + c_1(1 - c_{i-1})$, $i = 2, \dots, n$ (式中、 c_i は i 個のエンジンの検索範囲であり、 c_1 は1個のエンジンの範囲である)) という仮定と等価である。

【0123】検討すべき重要なバイアスが多くある。検索エンジンは、通常、検索フォームの背後に隠れた索引ドキュメントや、(ウェブ内検索用の) ロボットの除外基準や認証要件のために検索エンジンが除外されることとなるドキュメントのインデックス付けを検討しない。したがって本発明者らは、ウェブの真のサイズは、ここで推定されるものよりはるかに大きいと予想している。しかしながら、検索エンジンはこれらのドキュメントのインデックス付けを開始する可能性は低いことから、それら検索エンジンがインデックス付けを検討するウェブ (以下、「インデックス可能ウェブ」と称する) の大きさと、検索エンジンの相対的総合性を推定することは興味深いものである。

【0124】(a) 各検索エンジンがインデックス付けするウェブの量は、検索エンジン間でかなり変動するものであること、(b) 検索エンジンは独立にウェブのサンプリングを行わないことから、上記の対数関数による外挿は、インデックス可能ウェブのサイズを決定するには正確ではない。6個の検索エンジンのいずれも、ユーザー

がそのページを登録できる登録機能を提供する。多くのユーザーが、これら検索エンジンのいくつかでページを登録すると考えるのが妥当である。したがって、各検索エンジンがインデックス付けするページは、部分的に依存性である。各検索エンジンが行うサンプリング間の依存性の第2の原因は、各検索エンジンが、他のページにリンクされているページ、すなわちより人気のあるページの方に偏っているのが普通であるという事実から生じるものである。

【0125】図31における検索エンジンaと検索エンジンbの間の重複について考える。各検索エンジンがウェブを独立にサンプリングすると仮定すると、量 n_a/n_b (n_a は両方の検索エンジンによって返ってくるドキュメント数であり、 n_b は検索エンジンbによって返ってくるドキュメント数である) は、検索エンジンaによって網羅されるインデックス可能ウェブの割合 p_a の推定値である。基準点として6個の検索エンジンの検索範囲を用いると、 $p'_a = n_a/n_b$ と書くことができる (n_a は検索エンジンaによって返ってくる固有(unique)のドキュメント数であり、 n_b は6個の検索エンジンの組み合わせによって返って固有のドキュメント数である)。従って、 p'_a は、6個のエンジンの検索範囲に対する検索エンジンaの検索範囲であり、 $c = p'_a/p_b = n_a n_b / n_b n_b$ と書くことができる。この式を用いて本発明者らは、ここで検討している6個の検索エンジンによって網羅されるウェブの量に関連して、ウェブのサイズを推算する。検索エンジンのサイズはかなり変動することから、本発明者らは、最も小さい2つから最も大きい2つまでの2つの検索エンジンの組み合わせを用いて、 c の値を推定することを検討する。本発明者らはこの解析を、50以上のドキュメントを返す245の質問に限定する ($n_b = 0$ の場合の問題を回避するため)。表2にはその結果を示してある。 c の値が1より小さいと、インデックス可能ウェブのサイズが、6個のエンジン全てから検索されたドキュメントの数より小さいことが示唆される。相対的に大きいエンジンは、a) ユーザーが登録するページ以外のページを相対的に多くインデックス付けでき、b) ウェブ上の人気の低いページを相対的に多くインデックス付けできることから、依存性が相対的に小さいと予想できる。実際、検索エンジンが大きくなるにつれて c の推定値が高くなるという、明らかな傾向がある。

【0126】

【表2】

検索エンジン	Lycos と Infoseek	Infoseek と AltaVista	AltaVista と Northern Light	Northern Light と Excite	Excite と HotBot
エンジンの サイズ	最小		→		最大
推定 c	0.6	0.9	0.9	1.9	2.2
95% 信頼区間	±0.04	±0.06	±0.04	±0.12	±0.17

検索エンジン	応答時間の中央値(秒)
AltaVista	0.9
Infoseek	1.3
HotBot	2.6
Excite	5.2
Lycos	2.8
Northern Light	7.5
全検索エンジン	2.7
8 個検索エンジン中、最初の 検索エンジン	0.8
本発明のメタ検索エンジン からの最初の結果	1.3

最大の2つのエンジンでの比較からの $c = 2.2$ を用いて、エンジンが網羅するインデックス可能ウェブの割合を推算することができ、HotBotで17.8%、Exciteで14.1%、Northern Lightで13.8%、AltaVistaで13.3%、Infoseekで8.1%、Lycosで5.5%である。これらの結果を、図32の120に示してある。主要検索エンジンによってインデックス付けされたインデックス可能ウェブのパーセントは、一般に考えられているものよりかなり低い。(a) c の真の値は、2つの最も大きいエンジン間にある依存性のために、実際には2.2より大きいと考えられ、(b)異なる種類のユーザーからの質問については、異なる結果が認められる場合がある、ということは注目すべき点である。

【0127】報告によると、HotBotには5400万のページがあり、これは、インデックス可能ウェブのサイズについての下限に関する本発明者らの推定値を約3億万ページとするものである。Internet Archiveは、画像、音声などを除いて、8000万ページという推定値を用いている (Cunningham, M., "Brewster's millions", <http://www.irish-times.com/irish-times/paper/1997/0127/cmpl.html>, (1997))。Forrester Researchは、7500万を超えるページがあると推定している (Guilielmo, C., "Mr. Kurmit's neighborhood", Upside September, (1997))。AltaVistaは、現在、ウェブには1億~1億5000万ページがあると推定している (Brake, D., "Lost in cyberspace", New Scientist 154(2088), 12-13(1997))。

【0128】ページ検索時間の簡単な解析から、いくつか興味深い結論が得られる。表3には、6個の主要検索エンジンのそれぞれが応答するのに要した時間の中央値を、6個の検索エンジン全てに対して質問を同時に行った場合 (メタ検索エンジンで起こるような状態) に、これら検索エンジンのうち最初に結果を返したもののについて

【0129】

【表3】

これら各検索エンジンの応答時間および6個のエンジン中の最初のものの応答時間のヒストグラムを図33及び図34に示してあり、中央値を図35に示してある。図36には、 n の値を変えたときの n 個の検索エンジンのうちに最初に応答したものの応答時間の中央値を示してある。これらの結果は1997年9月のものであり、検索エンジンの相対的速度は、時間の経過に応じて変動する点は留意すべきである。

【0130】ここで、任意のウェブページをダウンロードするのに要する時間について見ると、図37に応答時間のヒストグラムがある。図38には、 n の値を変えたときの n 個の検索エンジン中最速のものが応答するのに要する時間の中央値を示してある。本発明のメタ検索エンジンが最初の結果を表示するまでに要する時間を推算することができる。この推算値は、それは6個の検索エンジンのうちの最初に応答するもの (このメタ検索エンジンは、実際には6個を超える検索エンジンを用いるが、本発明者らは、この場合は、主要なウェブ検索エンジンに絞っている) 及び10個のウェブページ (実際の数字は、応答する最初の検索エンジンが返す数によって決まる) の最初に表示されたものについての分布からサンプリングを行い、それらを一緒に加算し、1000回の試行について平均を取ることで得られる値である。

【0131】図39は、その分布のヒストグラムを示してある。分布の中央値は1.3秒である (それに対して、実際のページのダウンロードを行わない場合であっても、検索エンジンの応答時間中央値は2.7秒である)。比較のため、MetaCrawlerが結果を返すのに要する平均時間は25.7秒 (ページ確認を行わない場合) または139.3秒 (ページ確認を行う場合) である (Selberg, E. and Etzioni, O., "Multiservice search and comparison using the MetaCrawler", Proceedings of the 1995 World Wide Web Conference, (1995

5)。なお、基礎となる検索エンジン及び／またはウェブは、セルバーグとエチオニが実験を行った時より、かなり速いように思われる。

【0132】したがって、平均すると、本発明のメタ検索エンジンは、並列アーキテクチャであることによって、標準的な検索エンジンが結果を与えるより迅速に、最初のページを検出、ダウンロードおよび解析を行うことができることが分かる。ただし、標準的な検索エンジンは、ページのダウンロードや解析は行わない。なお、ここで述べた結果は、実行した特定の質問（質問の関数としての速度は検索エンジンごとに異なる）とそれを行ったネットワーク条件に固有のものであることに、留意すべきである。これらの要素は、結果をある特定の検索エンジンに偏らせる場合がある。ウェブのアクセス時間の非定常性はここでは考慮しない。例えば、検索エンジンの速度は、時間経過に伴ってかなり変動する。短期的変動は、ネットワークもしくは機械の問題およびユーザーのロード動作によるものであると考えられ、長期的変動は、検索エンジンのソフトウェア、検索エンジンのハードウェア資源または関連するネットワークの接続における変化によるものであると考えられる。

【0133】本発明のメタ検索エンジンは、ウェブ検索エンジンから返されるドキュメントのリアルタイム解析が可能であることを示している。実際、ウェブ検索エンジンの呼び出しおよびウェブページのダウンロードを並行して行うことで、本発明のメタ検索エンジンは、平均すると、標準的な検索エンジンを用いた場合より迅速に、最初の結果を表示することができる。

【0134】ユーザーからのフィードバックは、質問用語周囲のリアルタイムでのローカル文脈の表示とビュー時の文書中での質問用語の強調によって、ウェブ検索の効率が大幅に向上することを示している。

【0135】本発明者らの実験は、主要な検索エンジンの検索範囲についての上限が、インデックス可能ウェブの6% (Lycos) から18% (HotBot) までの間で変動することを示している。6個の検索エンジンの結果を合わせると、検索エンジン1個のみを用いた場合と比較して、3.5倍を超えるドキュメントが返ってくる。検索エンジン間の重複を解析することによって、本発明者らは、インデックス可能ウェブのサイズについての大体の下限は3億ページであると推定する。主要な検索エンジンが返す無効リンクの割合は、3%~7%の範囲で変動する。本発明者らの結果からは、主要なウェブ検索エンジンの相対的な検索範囲が分かり、セルバーグ及びエチオニが示しているように、いずれか1個の検索エンジンの範囲は非常に限られたものであることが確認される。

【0136】

【発明の効果】以上説明したように本発明は、質問をインターネット上にある複数のサードパーティ画像検索エンジンに送り、サードパーティ画像検索エンジンからの

応答を解析することで、質問に合致する画像に関する情報を抽出し、質問に合致する画像をダウンロードし、画像のサムネイル画像をユーザーに表示することにより、メタ検索の効率が向上して、その利用者が真に要求している画像データを迅速かつ的確にその利用者に提示できるようになるという効果がある。

【図面の簡単な説明】

【図1】本発明の好ましい実施の形態のメタ検索エンジン（ここでは、NEC Iメタ検索エンジン）のホームページを示す図である。

【図2】本発明の好ましい実施の形態のメタ検索エンジンのオプションページを示す図である。

【図3】本発明の好ましい実施の形態のメタ検索エンジンにおける「"nec"及び"digital watermark (デジタル電子透かし)"」という質問に対する応答例の第1の部分を示す図である。

【図4】本発明の好ましい実施の形態のメタ検索エンジンにおける「"nec"及び"digital watermark"」という質問に対する応答例の第2の部分を示す図であり、用語近接度情報を含む関連性尺度によってページをランク付けした結果を示している。

【図5】本発明の好ましい実施の形態のメタ検索エンジンにおける「"nec"及び"digital watermark"」という質問に対する応答例の第3の部分を示す図であり、質問用語のうちの1つのみが見つかったページを示している。

【図6】本発明の好ましい実施の形態のメタ検索エンジンにおける「"nec"及び"digital watermark"」という質問に対する応答例の第4の部分を示す図であり、質問用語が見つからなかったページを示している。

【図7】本発明の好ましい実施の形態のメタ検索エンジンにおける「"nec"及び"digital watermark"」という質問に対する応答例の第5の部分を示す図であり、先に見つけたページに対する重複文脈文字列を含むページを示している。

【図8】本発明の好ましい実施の形態のメタ検索エンジンにおける「"nec"及び"digital watermark"」という質問に対する応答例の第6の部分（最後の部分）を示す図であり、個々の独立した検索エンジン等から得られた結果を含む要約情報を示している。

【図9】本発明の好ましい実施の形態のメタ検索エンジンのページ表示の例を示す図であって、質問用語をハイライト表示するとともに、先頭部にあるリンクの表示によって、各質問用語がそれぞれ最初に現われたところからジャンプすることができることを示している。

【図10】本発明の好ましい実施の形態のメタ検索エンジンの概略制御フロー図であって、ページ検索デーモンとの相互作用を破線で示す図である。

【図11】本発明の好ましい実施の形態での画像のメタ検索を説明する概略制御フロー図であって、ページ検索デーモンとの相互作用を破線で示す図である。

【図12】写真についてフィルタをかけた画像データベースにおける「"koala(コアラ)"」という質問に対する、本発明の好ましい実施の形態のメタ検索エンジンの応答例の第1の部分を示す図である。

【図13】写真についてフィルタをかけた画像データベースにおける「"koala"」という質問に対する、本発明の好ましい実施の形態のメタ検索エンジンの応答例の第2の部分を示す図である。

【図14】グラフィックについてフィルタをかけた画像データベースにおける「"koala"」という質問に対する、本発明の好ましい実施の形態のメタ検索エンジンの応答例を示す図である。

【図15】質問「joydeep qhosh」についてのクラスターを示す図である。

【図16】図16には、質問「joydeep qhosh」についての最初の2つのクラスターの概要を示す図である。

【図17】質問「joydeep qhosh」に対する検索エンジンHuskySearchからのクラスターの第1の部分を示す図である。

【図18】質問「joydeep qhosh」に対する検索エンジンHuskySearchからのクラスターの第2の部分を示す図である。

【図19】質問「joydeep qhosh」に対する検索エンジンAltaVistaからのクラスターを示す図である。

【図20】質問「neural network」に対して、本発明の好ましい実施の形態のメタ検索エンジンによって得られたクラスターを示す図である。

【図21】質問「"Typing"及び"Injury"」に対して、本発明の好ましい実施の形態のメタ検索エンジンで得られるクラスターと、最初のクラスターの概要とを示す図である。

【図22】「NASDAQは何を表すか(What does NASDAQ stand for)」という質問についての、本発明の好ましい実施の形態のメタ検索エンジンの応答を示す図である。

【図23】「NASDAQは何を表すか」という質問についての、検索エンジンInfoseekの応答を示す図である。

【図24】「虹はどのようにして作られるか(How is a rainbow created?)」という質問についての、本発明の好ましい実施の形態のメタ検索エンジンの応答を示す図である。

【図25】「虹はどのようにして作られるか」という質問についての、検索エンジンInfoseekの応答を示す図である。

【図26】「ミーリー機械とは何か(What is a Mealy machine?)」という質問についての、本発明の好ましい実施の形態のメタ検索エンジンの応答を示す図である。

【図27】ある質問についての新たなヒットと、最近変更されたURL(uniform resource locator)とを示すホームページの例を示す図である。

【図28】最後に閲覧したとき以降にページに対して加

えられたテキストを示す、ページ表示の例を示す図である。

【図29】500回の質問に対して平均した、6つの検索エンジンのそれぞれの検索範囲を、これらの検索エンジン全体から得られる検索範囲と対比して示すグラフである。

【図30】検索エンジン数を増やした場合に、500回の質問に対する平均として、全体としての検索範囲の変化を示すグラフであって、図中の指数曲線は、検索エンジン数に対して対数的に検索範囲が増大すると仮定したときの曲線であり、検索エンジン数が増えるとともに、より多くのドキュメントが検索結果として与えられることがはっきりと分かる。

【図31】検索エンジンでは考慮されないウェブを除いたインデックス付け可能なウェブの数を推定するために、6つの検索エンジンの全体から結果として得られたドキュメントの数と、検索エンジン間で重複する数とを比較した結果を示す図である。

【図32】インデックス付け可能なウェブの推定数(この推定数は、真の値より小さいと予想されている)に対する、各検索エンジンの検索範囲を示すグラフである。

【図33】(a)~(c)は、それぞれ、主要な検索エンジンの応答時間のヒストグラムであり、これらヒストグラムにおいて、頻度は、ヒストグラムの各区間ごとにその区間内の応答時間の比率を示すように、正規化されている。

【図34】(a)~(c)は、それぞれ、主要な検索エンジンの応答時間のヒストグラムであり、(d)は、6個のエンジンに対して同時に質問を行った場合の最初の応答についての、過去の分布から導かれた10000サンプルから求めた応答時間のヒストグラムであり、これらヒストグラムにおいて、頻度は、ヒストグラムの各区間ごとにその区間内の応答時間の比率を示すように、正規化されている。

【図35】ウェブ検索エンジンごとに、その検索エンジンが応答するまでの時間の中央値を示すグラフである。

【図36】n個のウェブ検索エンジンの中で最先に応答したものについて、その応答が得られるまでに要した時間の中央値をnの値ごとに示すグラフである。

【図37】任意のウェブページについての応答時間の分布を示すヒストグラムである。

【図38】同時に要求したn個のページの中で最先に応答したものについて、それをダウンロードするのに要した時間の中央値をnの値ごとに示すグラフである。

【図39】メタ検索エンジンが最初の結果を表示するのに要する応答時間を示すグラフである。

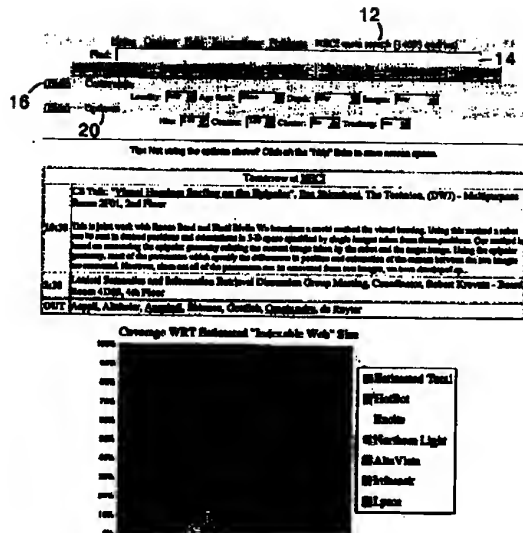
【符号の説明】

201 並列ページ検索エンジン

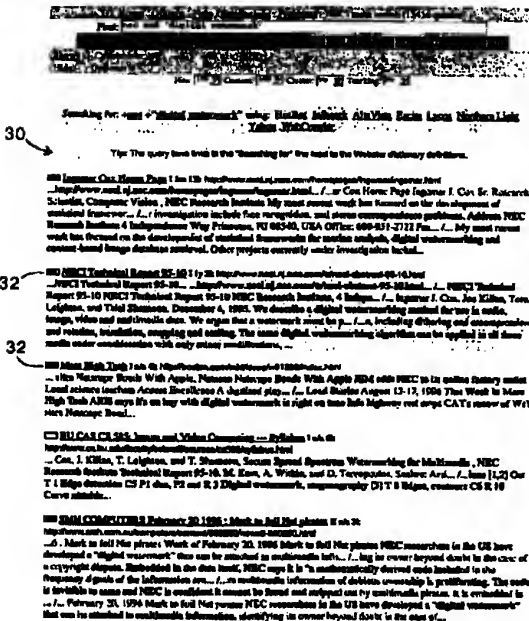
202 ウェブ

211~225, 231~247 ステップ

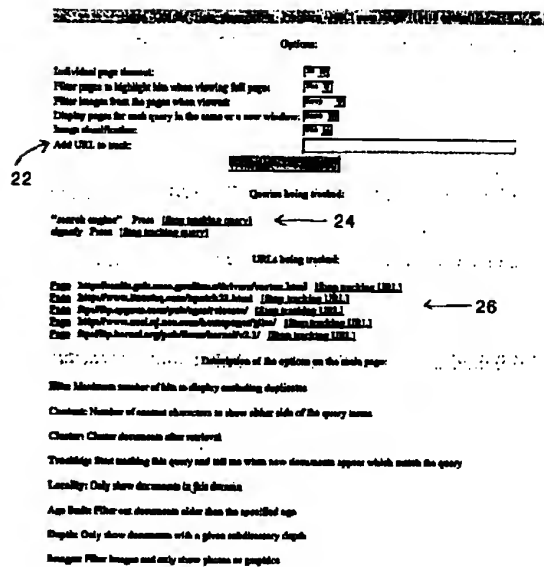
【図1】



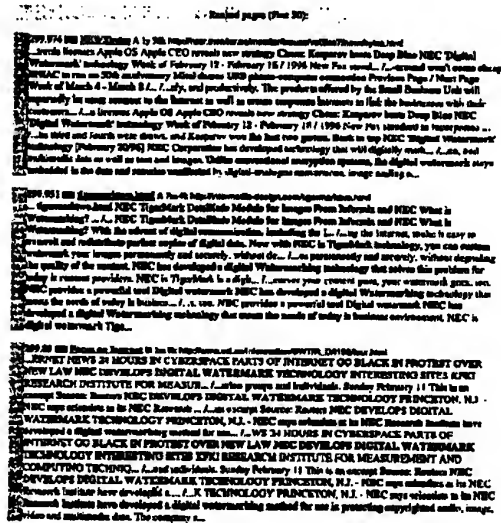
【図3】



【図2】



【図4】



【図5】

Only 1 search term was found in these documents:

- ARIS Technologies' Homepage H 16d 2k <http://www.musicode.com/welcome.html>
 ... ARIS Technologies' Homepage ARIS Technologies is an industry leader in digital watermarking. We deal exclusively with protecting intellectual property such as audio, video, and multimedia...
- Psych 267 Final Projects A n/a 8k <http://white.stanford.edu/~hoefer/psych267/final.html>
 ... interactive lighting design. Proceedings Eurographics '95, p. 229-240, 1995 (preprint). Digital watermark.
 References: Cox, Kilian, Leighton and Shamoon, "A Secure, Imperceptible yet Perceptually sal... /... IBM Tech.
 Report (preprint available). Further links to other papers and resources on digital watermarks. Face recognition
 with "eigenfaces". References: Turk and Pentland, "Face recognition u...
- Digimarc receives funding of \$4.5M. N n/a 6k <http://www.nisearch.com/cgi-bin/pdserv.pl?cbredd=YY19970425030163058&ho=typhoon&po=5005>
 ... Summary: First licensee is Adobe. Digimarc, the company that last year announced its Imagemarc digital
 watermark technology, seems to be ready to make its move in the market. ...
- Newsbytes Daily Summary N ** tid ** 25k <http://newsbytes.mpx.com.au/newsbytes/daily.html>
 ... Lemout & Hauspie [NASDAQ:LHSPF] (L&H) of Burlington, Massachusetts, and Ieper, Belgium. CHIPS NEC
 Develops World's Smallest Transistor TOKYO, JAPAN, 1997 SEP 11 (NB) - By Martyn Williams. NEC Co... /... PS
 NEC Develops World's Smallest Transistor TOKYO, JAPAN, 1997 SEP 11 (NB) - By Martyn Williams. NEC
 Corporation [TOKYO:6701] says it has developed the world's smallest operational transistor, a me... /... length of
 14 nanometers (14 millionths of a millimeter). The achievement was reached as part of NEC's development of a 10
 terabit memory chip. Intel Advances Mobile PC Platform HONG KONG, CHINA...
- A letter from the publisher TIME, December 6, 1971. L 1y 3k
<http://electron.rutgers.edu/~myadav/war71/wall/doc28.html>
 ...tough warning to India. But the only evidence of war that night was the blackout which was quite unnecessary."
 From the correspondents' files, and from background research assembled by Reporter-Res...

【図6】

No search terms were found in these documents:

- Article Two H ty 1k <http://miavx1.muhio.edu/~whitja/articletwo.html>
 Userdir rule failure The server was unable to resolve the requested / username reference, possible causes include:
 Username invalid Server is unable to determine username's login directory due to insufficient privilege to
- Jonathan G. Campbell, University of Ulster, N. Ireland, WWW Links t n/a 1k <http://www.ison.ulst.ac.uk/~jon/book/>
 J.G. Campbell's Bookmarks From 27 August 1997 this page is "permanently" relocated to
<http://www.infml.ulst.ac.uk/jgc/book> Updated 27 August 1997 - JG.Campbell@ulst.ac.uk
- CIO/Conserve WWW server address has changed L 5m 1k
<http://cioa.le.npi.edu:4997/mailboxes/comgrade06085104.118>
 CIO/Conserve WWW server address has changed The CIO web server address has changed. It is now
<http://www.cio.org> Please note too the new email address for the Conserve email interface to
- DEFINE IMAGE L 9m 2k <http://fram.fr/doc/aic/node58.html>
 DEFINE IMAGE Next: DEFINE /LIKE Up: DEFINE Previous: DEFINE HEADER DEFINE IMAGE DEFINE
 IMAGE Var1 File1 Key1 [Var2 File2 Key2 [...]] [/GLOBAL]
- Arizona Off-Road L 3m 3k <http://www.azoffroad.com>
 Arizona Off-Road 1833 W. Mountain View Road Phoenix, AZ 85021 ATC's MOTORCYCLES JET SKIS GO
 CARTS
- Résultats dans les cantons L 21d 5k <http://www.admin.ch/ch/v/pore/va/19840226/can316.html>
 Votation no 316 - Résultats dans les cantons Tableau récapitulatif / deutsch Votation no 316 Résultats dans les
 cantons Arrêté fédéral concernant la perception d'une redevance sur le trafic des poids lourds du 24 juin 19
- "We Know How the Parisians Felt" L 1y 6k <http://electron.rutgers.edu/~myadav/war71/wall/doc27b.html>
 "We Know How the Parisians Felt" "We Know How the Parisians Felt" Section: Box Page, TIME, Dec. 27, 1971
 Time Correspondent Dan Coggan, who covered the war from Pakistani side, was in Dacca when that city
 surrendered. His report
- The U.S. : A Policy in Shambles L 1y 6k <http://electron.rutgers.edu/~myadav/war71/wall/doc20b.html>
 The U.S. : A Policy in Shambles The Nixon Administration drew a fusillade of criticism last week for its policy on
 India and Pakistan. Two weeks ago, when war broke out between two traditional enemies, a State Department
 spokesman issued
- ClariNet Tearsheet: Government, Business, and General News N ** 04 ** 3k <http://www.clarinet.net/Samples/nb-other.html>
 ClariNet Tearsheet: Government, Business, and General News ClariNet * ClariNet Tearsheet: Government,
 Business, and General News ClariNet Tearsheet: Government, Business, and General News This summary of
 computer and technology news is

【図7】

Pages with duplicate context strings to a page above:

Vol. 1, No. 19 A by 18k <http://www.media.sbxpos.com/BULL/BUL0119.HTM>
 Aicmate H by 18k <http://www.seyboldreport.com/BULL/BUL0119.HTM>
 ...ore than five million members Pressek floats additional stock Pearlsitters launched in Europe NEC announces
 digital watermark Oracle to include software suite with Internet box Ca... BT have launched Pressek's
 D. Pearlsitters in seven European countries. Reuters reports that NEC claims to have developed a digital watermark
 system that could protect digital files... n members Pressek floats additional stock Pearlsitters launched in
 Europe NEC announces digital watermark Oracle to include software suite with Internet box Canon combines
 divisions within an L... tters in seven European countries. Reuters reports that NEC claims to have developed a
 digital watermark system that could protect digital files, such as still images, video and audio, from unauth...

<http://www.videodiscovery.com/vdyweb/dvd/dvdfaq.txt> H 3m 118k
<http://www.videodiscovery.com/vdyweb/dvd/dvdfaq.txt>
 ...a on NTSC line 21. The digital standard (CGMS/D) is not yet finalized, but will apply to digital connections such
 as IEEE 1394/Firewire. 3) Because of the potential for perfect digital copies, paramod... /...isplaying it. No
 unscrambled digital output is allowed until work in progress for secure digital connections is finished. On the
 computer side, DVD-ROM drives and video display/decoder hardware or softw... /...d a PCM audio track. (Other
 streams such as Dolby Digital audio, MPEG audio, and subpicture are not necessary for the simplest case.) Basic
 DVD control codes are also needed. At the moment it's difficult... /...doing this, but it's possible. The music industry
 is also requesting an "embedding signalling" or "digital watermark" copy protection feature. This applies a digital
 signature to the audio in the form of supposedly ...

Hyflex J1 Launch H 3m 3k <http://jpn.co.jp/arc96/p14.html>
 Hyflex J1 Launch H 3m 3k <http://jpn.co.jp/arc96/p14.html>
 Hyflex J1 Launch NEC Develops Digital Watermarking Technique JPN Scientists at NEC Research Institut...
 ...inch NEC Develops Digital Watermarking Technique JPN Scientists at NEC Research Institute in Princeton,
 NJ, have developed a digital watermarking method that could be us... /...ary information is increasingly an issue,"
 said Tatsuo Ishigoro, associate senior vice president of NEC Corp. "I am convinced that our watermarking
 technique is a solution that will be welcomed espec... /... Hyflex J1 Launch NEC Develops Digital Watermarking
 Technique JPN Scientists at NEC Research Institute in Princeton, NJ, have de... /...ique JPN Scientists at NEC
 Research Institut in Princeton, NJ, have developed a digital watermarking method that could be used to protect
 the copyright of images and music on the Internet. Con... /...e is no way to track its reproduction and therefore it
 provides little protection against piracy. A digital watermark, however, can protect a copyright by means of an
 invisible identification code that is permanently...

Internet H 4/s 20k <http://nat.info.nl/ur0296/Internet.html>
 ...scape servers. Dit kwam o.a. door het feit dat bepaalde optionele onderdelen zoals een database-connector duur
 betaald moeten worden. Microsoft op zijn heurt deed daar weer een schepje bovenop door we...
 .../.../www.microsoft.com/infosec: <http://www.microsoft.com/windows> <http://www.netscape.com> NEC ontwikkelt
 Digital Watermark Technology NEC is in zijn computerlaboratoriums bezig met een digi...
 ...microsoft.com/windows <http://www.netscape.com> NEC ontwikkelt Digital Watermark Technology NEC is in
 zijn computerlaboratoriums bezig met een digitaal watermerk. Dit watermerk maakt in de toekomst... /...com/infosec
<http://www.microsoft.com/windows> <http://www.netscape.com> NEC ontwikkelt Digital Watermark Technology
 NEC is in zijn computerlaboratoriums bezig met een digitaal watermerk. Dit watermerk ...

【図13】

This search: [koolz](#) Search engine pages: [AltaVista Images](#) [Coral Images](#) [HotBot Images](#) [Page 2](#) [Page 3](#) [Page 4](#) [Page 5](#) [Page 6](#) [Lycos Images](#) [Page 2](#) [Page 3](#) [Page 4](#) [Page 5](#) [WebSeer](#) [Yahoo Images](#)

Engine	Response	Total	Retrieved	Processed	Duplicates
AltaVista Images	Yes	0	0	0	0
Coral Images	Yes	7	7	7	0
HotBot Images	Yes	5511	125	99	1
Lycos Images	Yes	222	80	85	1
WebSeer	Yes	0	0	0	0
Yahoo Images	Yes	4	4	4	1
Total		5744	216	195	3

More documents were found but the maximum number of hits was reached.

Filtered due to size: 12 Filtered due to type: 21

【図8】

These documents no longer exist

Error 404 Not found - file doesn't exist or is read protected (even tried multi) Digital Image Watermarking: Main Project Page <http://www.csuglab.com/edu/info/People/vbadar/cs631/wm/kproy/project.html>

Error 404 Not found Labeling Techniques for Multimedia Data: http://www.it.ei.tudelft.nl/pda/smash/public/benlux_cr.html

Error 404 Not found Labeling Techniques for Multimedia Data: http://www.it.ei.tudelft.nl/pda/smash/public/benb96/benlux_cr.html

Error 404 Not Found Artisoft Inc. -- Industry Awards and Recognition <http://artisoft.com/main/overview/awards.html>

Error 404 File Not Found The Rutgers Review <http://electron.rutgers.edu/~nebus/>

This search: **neg + "digital watermark"** Search engine pages: [AltaVista Page 2](#) [Page 3](#) [Excite Page 2](#) [HotBot Page 2](#) [Infoseek Lycos](#) [Northern Lights Page 2](#) [WebCrawler](#) [Yahoo](#)

Query expansion (adding these words to the query may help): **digitally** (16) **digitized** (16) **digit** (9) **digitale** (8) **digital** (8) **digitization** (5) **digits** (3) **digitize** (3) **watermarking** (463) **watermarks** (127) **watermarked** (50)

Engine	Response	Total	Retrieved	Processed	Duplicates
AltaVista	Yes	29	29	29	11
Excite	Yes	26	26	26	13
HotBot	Yes	39	39	33	5
Infoseek	Yes	19	19	17	0
Lycos	Yes	10	10	10	0
Northern Light	Yes	50	50	33	19
WebCrawler	Yes	1	1	0	0
Yahoo	Yes	0	0	0	0
Total		174	174	148	48

More documents were found but the maximum number of hits was reached.

2 terms: 70 1 term: 5 0 terms: 11 duplicate context: 9 invalid link: 5

【図9】

40



Jump to: [neg \(2\)](#) [digital watermark \(2\)](#) <http://www.neci.nj.nec.com/tr/neci-abstract-95-10.html> [Track page!](#)

■ NECI Technical Report 95-10

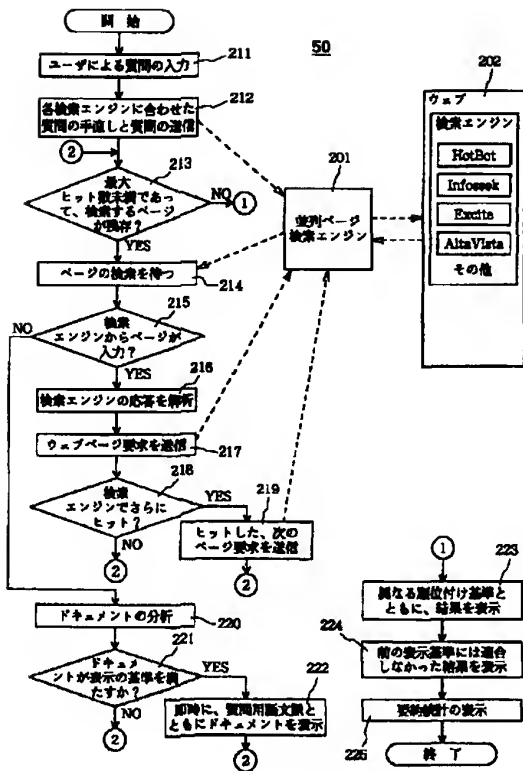
■ NEC Research Institute, 4 Independence Way, Princeton, NJ 08540.

Secure Spread Spectrum Watermarking for Multimedia

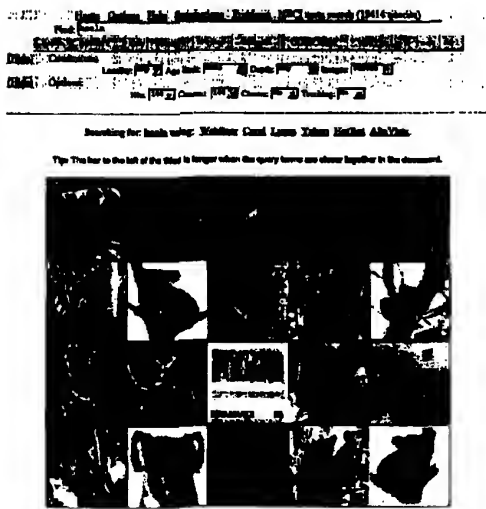
Ingeger J. Cox, Joe Kilian, Tom Leighton, and Talal Shamon. December 4, 1995.

We describe a  digital watermarking method for use in audio, image, video and multimedia data. We argue that a watermark must be placed in perceptually significant components of a signal if it is to be robust to common signal distortions and malicious attack. However, it is well known that modification of these components can lead to perceptual degradation of the signal. To avoid this, we propose to insert a watermark into the spectral components of the data using techniques analogous to spread spectrum communications, hiding a narrow band signal in a wideband channel that is the data. The watermark is difficult for an attacker to remove, even when several individuals conspire together with independently watermarked copies of the data. It is also robust to common signal and geometric distortions such as digital-to-analog and analog-to-digital conversion, resampling, and requantization, including dithering and recompression and rotation, translation, cropping and scaling. The same  digital watermarking algorithm can be applied to all three media under consideration with only minor modifications, making it especially appropriate for multimedia products. Retrieval of the watermark unambiguously identifies the owner, and the watermark can be constructed to make counterfeiting almost impossible. Experimental results are presented to support these claims.

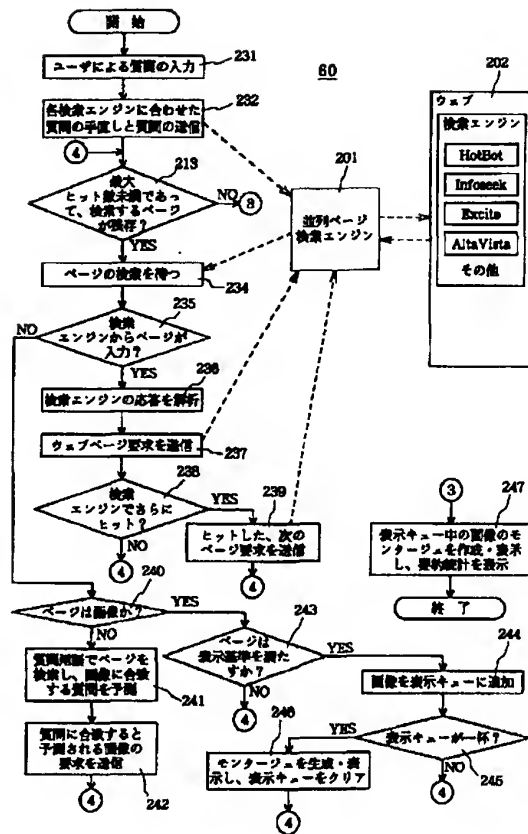
【図10】



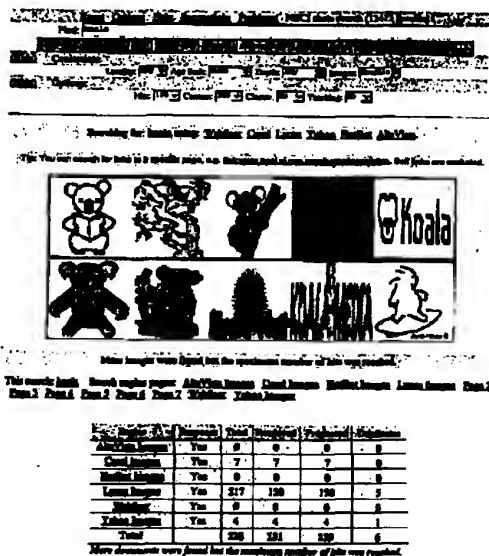
【図12】



【図11】



【図14】



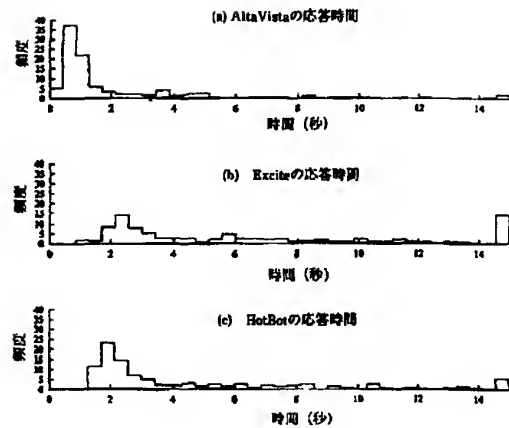
【図15】

Clusters:

70

university of texas
 artificial neural networks
 jose nelson amaral
 classifiers kagan tumor
 department of electrical
 combined neural classifiers
 applying genetic algorithms
 hybrid intelligent architecture OR hybrid intelligent architecture AND systems
 radial basis function
 classifier boundary distributions
 estimating the bayes
 boundaries in linearly
 pattern recognition
 tx 78712
 ksmall taha
 kagan plus
 international conference
 systems
 abstract
 paper
 austin
 ecc
 utexas

【図33】



【図17】



HUSKY SEARCH

Query: (joydeep ghosh)

76

Documents: 102, Clusters: 14, Average Cluster Size: 11.21 documents

Document Group	Size	Phrase and Sample Document Titles
Cluster 1	19	Artificial (74%), Hybrid Intelligent (32%), domain knowledge (26%) ORGANIZING COMMITTEE Unlabeled Abstract Abstract Unlabeled
Cluster 2	7	Click on to view the abstract and on to obtain a postscript copy-- (71%), postscript copy of the full paper (100%), paper is currently not available. (57%) Kagan-Tumors Publications Classification Machine Design Refined Archival Journal Publications (Full/Regular Papers) Technical Research
Cluster 3	15	Kagan tumor (100%), NEURAL CLASSIFIERS Kagan Tumor and Joydeep ghosh (27%), pattern classifiers (20%), tumor (53%) Unlabeled Unlabeled Abstract Abstract Abstract
Cluster 4	8	of Electrical and Computer (38%) Unlabeled Unlabeled LANE Home Page Joydeep Ghosh LANE Home Page
Cluster 5	31	area (35%), RESEARCH & EDUCATIONAL RESOURCES/ ORGANIZATIONS/ centers (10%), of Texas (10%), or (21%) Unlabeled Unlabeled Eaton 77 Chemical Engineering LANE Home Page Joydeep Ghosh
Cluster 6	12	University of Texas at austin (50%), University of Texas (67%), Texas (100%) Unlabeled Unlabeled LANE Home Page Joydeep Ghosh Joydeep Ghosh
Cluster 7	3	area Journal (100%) LANE Home Page Refined Archival Journal Publications (Full/Regular Papers) Refined Archival Journal Publications (Full/Regular Papers)

【図16】

Cluster summaries:

Document: ...by clicking on Journal Papers: Ismail Taha and Joydeep Ghosh, "A Hybrid Intelligent Architecture and Its Application to Water Reservo... /... Journal of Smart Engineering Systems, Ismail Taha and Joydeep Ghosh, "Symbolic Interpretation of Artificial Neural Networks", submitted ... /... Austin, 1996. Conference Papers: Ismail Taha and Joydeep Ghosh, "Evaluation and Ordering of Rules Extracted from Feedforward Networks... /... Also, Tech. Rep. TR-97-01-106, The Computer and Vision Research Center, University of Texas, Austin, 1996.

Conference Papers: Ismail Taha an...
Document: Joydeep Ghosh... /... Joydeep Ghosh Telephone: (512) 471-8980 Fax: (512) 471-5... /... Joydeep Ghosh Joydeep Ghosh Telephone: (512) 471-8980 Fax: (512) 471-5532 E-mail: ghosh@pine... /... Fax: (512) 471-5532 E-mail: ghosh@pine.cce.utexas.edu Address: The University of Texas at Austin Department of Electrical & Computer Engineering...

Document: Yoon Shin and Joydeep Ghosh Department of Electrical and Computer... /... Yoon Shin and Joydeep Ghosh Department of Electrical and Computer Engineering The University of Texas... /... in and Joydeep Ghosh Department of Electrical and Computer Engineering The University of Texas at Austin Austin, TX 78712 Abstract: This paper introduces a nov...
...more...

artificial neural networks

Document: Artificial Neural Networks Authors: Bryan W. Stiles and Joydeep Ghosh Department of Electrical and Computer Engineering The Univ... /... nity of Texas at Austin Correspondence: Bryan Stiles c/o Joydeep Ghosh Department of Electrical and Computer Engineering The Univ... /... Phone: (512) 471-2358 Email: bstiles@pine.cce.utexas.edu Joydeep Ghosh Department of Electrical and Computer Engineering The Univ... /... A Habituation Based Mechanism for Encoding Temporal Information in Artificial Neural Networks Authors: Bryan W. Stiles and Joydeep Ghosh Department o... /... ghosh@pine.cce.utexas.edu Submit to: Applications and Science of Artificial Neural Networks Steven K. Rogers and Dennis W. Ruck at AeroSense '9...

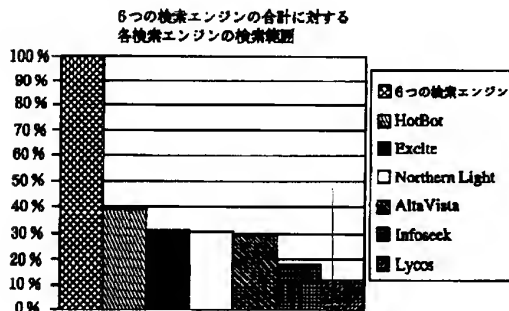
Document: (eds.), IEEE Press, 1995, pp. 135 - 144, Bryan W. Stiles and Joydeep Ghosh, "A Habituation Based Mechanism for Encoding Temporal Information in Arti... /... E Proc. Vol. , Orlando, April 1995, pp. Bryan W. Stiles and Joydeep Ghosh, "Habituation Based Neural Classifiers for Spatio-temporal Signals", Proc... /... Proc. ICASSP-95, Detroit, May 1995, pp. Bryan W. Stiles and Joydeep Ghosh, "Dynamic Neural Networks for the Classification of Oceanographic Data", ... /... Ghosh, "A Habituation Based Mechanism for Encoding Temporal Information in Artificial Neural Networks", (invited paper) Proc. SPIE Conf. on Applications and Science of Artif... /... tworks", (invited paper) Proc. SPIE Conf. on Applications and Science of Artificial Neural Networks IV, SPIE Proc. Vol. , Orlando, April 1995, pp. Bryan W. St...

Document: ...uth.edu Larry D. Jackel Robert E. Schapire Y. Freund Kagan Turner and Joydeep Ghosh Shimon Edelmann Jonathan Baxter Anders Krogh and Jesper Vedelsby ... /... from "ftp://eris.wisconsin.weizmann.ac.il/pub/mum.ps.Z" Kagan Turner and Joydeep Ghosh, "Theoretical Foundations of Linear and Order Statistics Combiners for Ne... /... When Networks Disagree: Ensemble Methods for Neural Networks", Chapter 10, Artificial Neural Networks for Speech and Vision, editor R.J. Mainwaring, Chapman-Hall, London 1993 M...

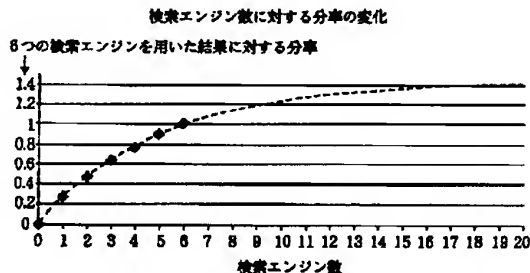
...more...

74

【図29】



【図30】



【図18】

80

Document Group	Size	Phrase and Sample Document Titles
Cluster 8	16	papers (100%) Kagan Tumer's Publications Classification Kurt's Publications Working Papers Refereed Archival Journal Publications (Full/Regular Papers)
Cluster 9	2	oms. Books (100%) LANS Home Page Books/Proceedings (Edited)
Cluster 10	8	(100%) LANS Home Page Associated Members LANS Home Page Associated Members No Title
Cluster 11	4	s, Austin (100%) Untitled LANS Home Page LANS Home Page No Title
Cluster 12	11	combining (100%), outputs (64%) Abstract Abstract Abstract Untitled Abstract
Cluster 13	8	ted (100%) LANS Home Page Kagan Tumer's Publications CIS Publications Database Kurt's Publications Refereed Archival Journal Publications (Full/Regular Papers)
Cluster 14	13	e (100%) LANS Home Page Jordsep Ghosh JOYDEEP GHOSH Jordsep Ghosh LANS Home Page
Unclassified	3	Musical Events in Mumbai Batch-Wise Mailing Lists and Home Pages for IITK Publications de José Nelson Amaral

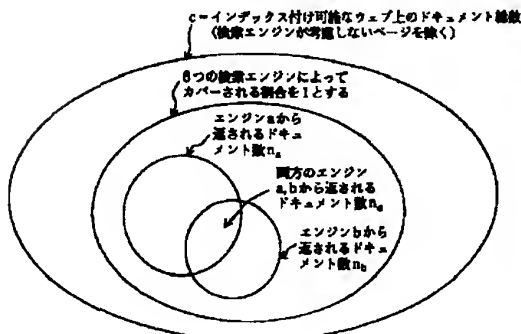
【図20】

Clusters:

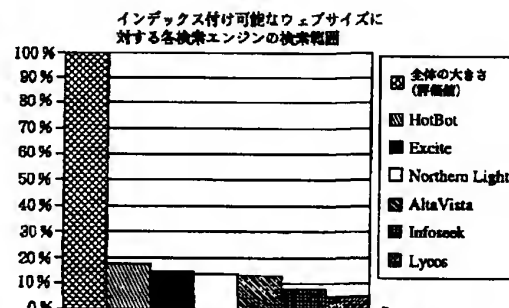
artificial neural networks
networks neural networks
ieee international conference
university of texas
pacific northwest laboratory
recurrent neural networks
nsw in hep
pacific northwest national
university of california
austrian research institute
self organizing map
northwest national laboratory
artificial intelligence
pattern recognition
research group
international conference
fuzzy logic
san diego
genetic algorithms
ieee transactions
signal processing
pacific northwest
technical report
machine learning
data
nets
ai
software

84

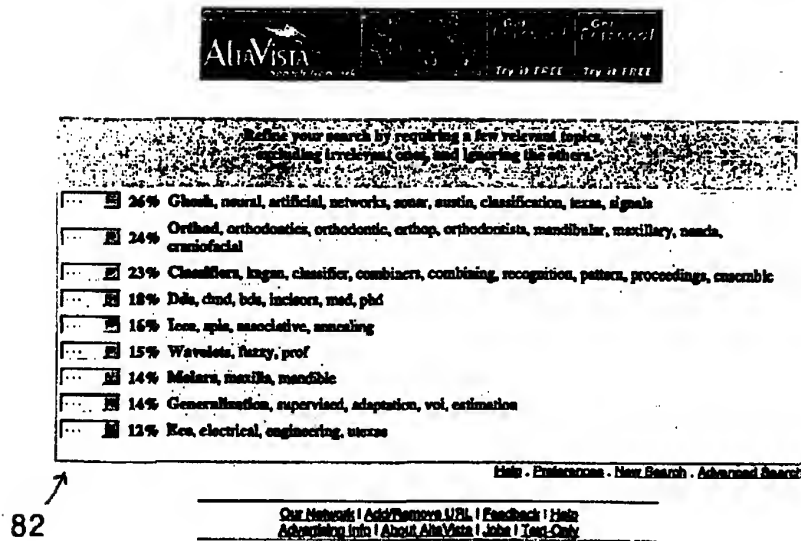
【図31】



【図32】



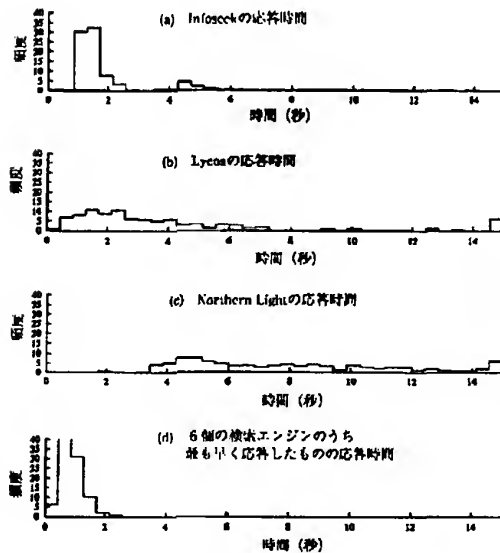
【図19】



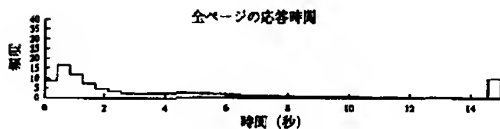
82

Digital Equipment Corporation
 Division / Private Network
 Copyright 1997 © All Rights Reserved

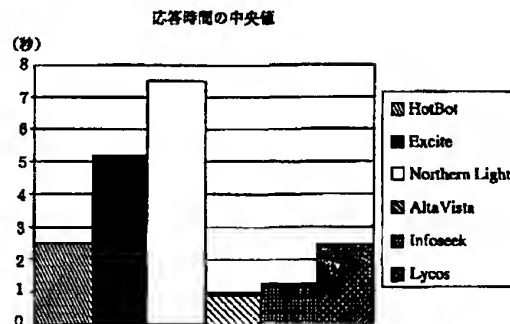
【図34】



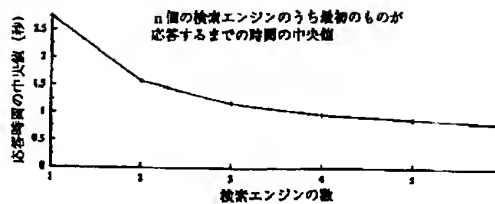
【図37】



【図35】



【図36】



【図21】

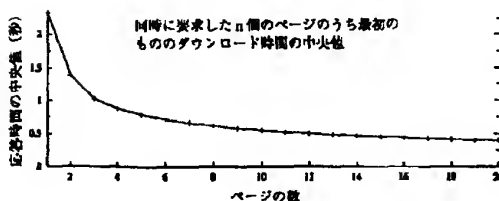
Cluster:

carpal tunnel syndrome
 repetitive stress injuries
 software monitoring tools
 contain useful advice
 repetitive strain injuries
 san jose state OR san jose state AND archive
 kinds of documents
 university of utah
 htm diana carroll
 products include split
 dan wallach
 keyboard alternatives
 interest finder OR browse groups
 human factors
 tifaq general
 injuries
 related
 archive
 resources
 contain useful advice AND resources OR contain useful advice AND keyboard alternatives
 repetitive stress injuries AND keyboard alternatives
 repetitive stress injuries AND dan wallach

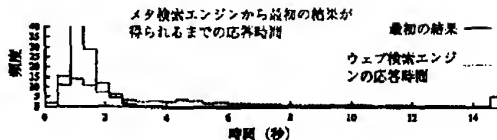
86

Document...FAQ - Typing Injury .../... Typing Injury FAQ Home Page (TIFAQ) [General] [Keyboards] [Speech] [Mice] [Software] ...FAQ - Typing Injury .../... Typing Injury FAQ Home Page (TIFAQ) [General] [Keyboards] [Speech] [Mice] [Software] .../... Injury Archive. sources of information for people with typing injuries, repetitive stress injuries, carpal tunnel syndrome, etc. The TIFAQ is targeted at computer users suffering at the hands of their equipment. You'll...
 Document...JOM's Ergonomics Page is a site by John Murray at University of Michigan that focuses on typing injuries, carpal tunnel syndrome and design concepts. Office Working Postures is a crime.../...eral links to various safety oriented covers, lists, and newsgroups. You do the searching. Typing Injuries is everything you ever wanted to know about typing injuries by Dan Wallach at Princeton.../... Princeton. Lots of publications and links. Everything you wanted to know and more. Typing Injury Archive is a typing injury library by Dan Wallach at Princeton. Here you will find a well categorized list of typing injury .../... and Ergonomics Home Page links to several ergonomic sites that focus on safety issues, such as: carpal tunnel syndrome, back injuries, air quality, sick building syndrome, and lighting. Lawrence Livermore Lab .../... technology and human factors engineering. Mostly in-house work but an interesting site. Carpal Tunnel Syndrome is a commercial site but does have lots of references to CTS. Specific emphasis is on keyboard...
 Document...at the keyboard. Site offers Time Out For Windows, an ergonomic exercise break program. Typing Injury FAQ: This is the home page for the Typing Injury FAQ and Typing Injury Archive. .../... an ergonomic exercise break program. Typing Injury FAQ: This is the home page for the Typing Injury FAQ and Typing Injury Archive. NEW! University of Minnesota Office Ergonomics.../...arc, how you get them, and some guidelines for how one may help heal oneself from this devastating injury. Carpal Tunnel Syndrome & Repetitive Stress Computer Related Repetitive Strain Injury: I hope on this page to provide a very brief introduction to RSI for the benefit of students who.../...em, and some guidelines for how one may help heal oneself from this devastating injury. Carpal Tunnel Syndrome & Repetitive Stress Computer Related Repetitive Strain Injury: I hope on this page to.../...page helps you avoid OOS / RSI with Micropause and Exercise Breaks. Patient's Guide to Carpal Tunnel Syndrome: The following documents attempt to explain what Carpal Tunnel Syndrome is, how it is diagnosed...
 Document...Typing Injury FAQ: General Information .../... Typing Injury FAQ: General Information General Information (TIFAQ) [General] [Keyboards] [Speech] [Mice] [Software] .../... Typing Injury FAQ: General Information General Information [TIFAQ] [General] [Keyboards] [Speech] [Mice] [Software] .../... mayon for RSI WRULD Work-Related Upper Limb Disorders - yet another synonym for RSI CTS Carpal Tunnel Syndrome (see below) Hyperextension Marked bending at a joint. Pronation Turning the palm down.../...over the wrist and forearm, more tenderness, and it gets worse with repetitive activity. Carpal Tunnel Syndrome the nerves that run through your wrist into your fingers get trapped by the inflamed aw...
 ...more...

【図38】



【図39】



【図22】

90

Home Options Help Suggestions Problems NECI meta search (14083 queries)

Find:

Search:

(Show) Constraints:

(Hide) Options:

Hits: Context: Cluster: Tracking:

Searching for: "NASDAQ stands for" "NASDAQ is an abbreviation" "NASDAQ means" using: HotBot
Infoseek AltaVista Excite Lycos Northern Light Yahoo WebCrawler.

Tip: For better precision with multiple terms you might like to use "+" to ensure that the results contain specific terms (e.g. "laser glasses" +optics).

Ref...viation for the New York Stock Exchange AMEX is an abbreviation for the American Stock Exchange
NASDAQ is an abbreviation for the National Association of Securities Dealers Automatic Quotation Exchange "Top 5% of the ...
Ref...nformation on NASDAQ and the companies traded thereon. (Incidentally, does anyone know what NASDAQ stands for?) NYSE All about the New York Stock Exchange. Data mongers too...
Ref... The NASDAQ Last-Updated: 25 Oct 1996 From: billmanr@aol.com , jeffwben@aol.com , cml@cs.umd.edu
NASDAQ is an abbreviation for the National Association of Securities Dealers Automated Quotation system. It is also common...
Ref... NASDAQ Last-Updated: 25 Oct 1996 From: billmanr@aol.com , jeffwben@aol.com , log@invest-faq.com
NASDAQ is an abbreviation for the National Association of Securities Dealers Automated Quotation system. It is also common...
Ref...ble for the operation and regulation of the NASDAQ stock market and over-the-counter markets. NASDAQ stands for the National Association of Securities Dealers Automated Quotation System. A nationwide computer...
Ref...site Index is a value weighted index that monitors more than 2,000 stocks traded over-the-counter. NASDAQ stands for National Association of Securities Dealers Automated Quotations. It has been available since 1971...
Ref...as an incentive stock option under Section 422 of the Code. (k) "NASDAQ" means the National Association of Securities Dealers, Inc. Automated Quotation System...

[...results deleted...]

This search: "NASDAQ stands for" "NASDAQ is an abbreviation" "NASDAQ means" Search engine pages:
AltaVista Excite HotBot Infoseek Lycos Northern Light WebCrawler Yahoo

Engine	Response	Total	Retrieved	Processed	Duplicates
AltaVista	Yes	9	9	9	1
Excite	Yes	24	24	24	1
HotBot	Yes	23	23	23	4
Infoseek	Yes	5	5	5	0
Lycos	Yes	0	0	0	0
Northern Light	Yes	0	0	0	0
WebCrawler	Yes	0	0	0	0
Yahoo	Yes	0	0	0	0
Total		61	61	61	6

3 terms: 0 2 terms: 0 1 term: 26 0 terms: 9 duplicate context: 14 cannot access: 3 invalid link: 3

[23]



92

Home | Add URL | Free Software | Help



Infoseek found 23,064,236 pages containing at least one of these words: what does NASDAQ stand for?

Related Topics:

- Stock Research NASDAQ - are you a stressed investor? FREE copy of the best
- Stock Research NASDAQ - are you a stressed investor? FREE copy of the best
- Market, Stock Research NASDAQ that always call it right. Our ...
- 72 % <http://www.applinet.com/stockresearch/> (Site 5.3K)

CyberLead BQ

CyberCorp designs and develops real-time decision support, execution and trading systems for NASDAQ stock market traders. Be sure to check out CyberTrade.

63 % <http://www.cyber-corp.com/> (Site 4.6K)

InvestQuest® AMUSEMENT & RECREATION SERVICES

AMUSEMENT & RECREATION SERVICES. [InvestQuest® Home Company List Industrial Company Search] ALLIANCE GAMING CORP (NASDAQ:ALLG) ALPHA HOSPITALITY CORP (NASDAQ:ALHY) AMERICAN ...

63 % http://www.investquest.com/stock/79_industry.htm (Site 11.8K)

InvestQuest® INSURANCE CARRIERS

INSURANCE CARRIERS. [InvestQuest® Home Company List Industrial Company Search] 20TH CENTURY INDUSTRIES (NYSE:TW) ACCEL INTERNATIONAL CORP (NASDAQ:ACIL) ACCEPTANCE INSURANCE ...

63 % http://www.investquest.com/stock/63_industry.htm (Site 26.8K)

InvestQuest® WHOLESALE TRADE-DURABLE GOODS

WHOLESALE TRADE-DURABLE GOODS. [InvestQuest® Home Company List Industrial Company Search] AAR CORP (NYSE:AAR) ALATIX ENVIRONMENTAL CORP (NASDAQ:ALUQ) ACE HARDWARE CORP ...

63 % http://www.investquest.com/stock/50_industry.htm (Site 34.2K)

Stocks by Symbol - C

Stocks by Symbol - C. C (NYSE) Chrysler CA (TSE) Canadian Airlines CANOTC/3AMQIS (OTC) Magisoft Software Corp. CAWS+ (NASDAQ) CAI Wireless Systems Inc. CBMI (NASDAQ) Creative ...

63 % <http://stockinfo.com/stock/symbol-c/index.html> (Site 9.6K)

Stocks by Company Name - A

A+ Communications (NASDAQ:AACOM) A. G. Edwards (NYSE:AGE) Alcatel Environmental (NASDAQ:ASDQ) ascena health (NASDAQ:ACSC) acslm (NASDAQ:AKLM) Actarkey Communications (NASDAQ:AK) ...

63 % <http://stockinfo.com/stock/company-a/index.html> (Site 8.5K)

InvestQuest® RUBBER AND MISC. PLASTICS PRODUCTS

RUBBER AND MISC. PLASTICS PRODUCTS. [InvestQuest® Home Company List Industrial Company Search] ADVANCED MATERIALS GROUP INC (NASDAQ:ADMG) AEP INDUSTRIES INC (NASDAQ:AEP) ...

63 % http://www.investquest.com/stock/30_industry.htm (Site 18.6K)

InvestQuest® FABRICATED METAL PRODUCTS

FABRICATED METAL PRODUCTS. [InvestQuest® Home Company List Industrial Company Search] AAYID THERMAL TECHNOLOGIES INC (NASDAQ:AATT) ABC RAIL PRODUCTS CORP (NASDAQ:ABRC) ABS ...

63 % http://www.investquest.com/stock/74_industry.htm (Site 13.3K)

InvestQuest® ENGINEERING & MANAGEMENT SERVICES

ENGINEERING & MANAGEMENT SERVICES. [InvestQuest® Home Company List Industrial Company Search] ADVANCED DETECTORS INC (OTC Bulletin Board:ADET) AERO SYSTEMS ENGINEERING INC ...

63 % http://www.investquest.com/stock/77_industry.htm (Site 15.8K)

More Summaries | next 36

Copyright © 1995-97 InfoSeek Corporation. All rights reserved.
 Infoseek incorporates Libantix technology from InfoLight.
 Disclaimer

【図24】

94

Home Options Help Suggestions Problems NECI meta search (13581 queries)

Find: how is a rainbow created?

Advanced Search Basic Search

(Show) Constraints (Hide) Options

Hits: 160 Context: 256 Cluster: No Tracking: No

Searching for: "rainbow is created" "makes a rainbow created" "rainbow is produced" "rainbow is made" using: HotBot Infoseek AltaVista Excite Lycos NorthernLight Yahoo WebCrawler

Tip: For better precision with multiple terms you might like to use "+" to ensure that the results contain specific terms (e.g. "the glass" +optics).

Ref:... the green flash, it helps to know how our atmosphere effects sunlight. Coincidentally, the phenomenon responsible for the green flash is also the one that paints rainbows across Hawaii's sky. A rainbow is created when rays of sunlight enter a raindrop, bounce around inside, and exit. Light from the sun consists of a pourpour of colors that are each bent by a different amount inside a raindrop. This unique...

Ref:...scapes the raindrop after it is reflected once. A part of the ray is reflected again and travels along inside the drop to emerge from the drop. The rainbow we normally see is called the primary rainbow and is produced by one internal reflection the secondary rainbow arises from two internal reflections and the rays exit the drop at an angle of 50 degrees rather than the 42degrees for the red primary bow. ...

Ref:...c rainbow we do not see the sun, and we rarely see a rainbow in winter. How do we explain this appearance of a bow, double bows, size of arc, and brightness of the rainbow? Answer The rainbow is produced by sunlight passing through a raindrop or a collection of rain drops. A typical raindrop is spherical and as a light ray strikes the surface of the raindrop, some light is reflected and some passes ...

Ref:...se to us. He promised that the earth will never be destroyed again by a flood. As a sign of that promise He put a rainbow in the sky. Whenever we see a rainbow, we can think of God's promise. The rainbow is made up of all the colors. Back To Index Next Page... Page 1 ...

Ref:...two rainbows, the narrower male rainbow and the wider female. The male rainbow can not stop the rain by itself. When it is followed by the female the rain stops. Other Native Americans believe the rainbow is made from the souls of wild flowers that lived in the forest and lilies from the prairies. A Japanese myth tells of the first man Isanagi and the first woman Isanami who stood on the floating bridge of amatsara before the clear light of Nirvana or heaven. In Arabia the rainbow is a tapestry draped by the hands of the south wind. It is also called the cloud's bow or Allah's bow. In Islam the rainbow is made up of four colors red, yellow, green and blue related to the four elements. In myths of India the Goddess Indra not only carries a thunderbolt like the Greek God Zeus but she also carries a ...

Ref:... true b. false 13. The average speed of light is greatest in __. a. red glass. b. orange glass. c. green glass. d. blue glass. e. is the same in all of these. 14. The secondary rainbow is produced with an extra (choose the best answer) a. dispersion. b. reflection. c. refraction. d. diffraction. 15. If a person has green cones that are weak, then yellow light will appear _ to ...

Ref:...ever wonder what makes the color in a rainbow ? The answer is sunlight, it has all of the colors of the rainbow in it, but they are all mixed up together so you are not able to see them. The rainbow is made up of drops of water. When sunlight passes through a drop of water, it bends and the colors inside the light split apart and are separated so that we can see them. When the sunlight passes through...

[...section deleted...]

【図26】

100

Home Options Help Suggestions Problems NECI meta search (13584 queries)

Find:

Advanced Search Basic Search Images Citations WebCrawler

(Show) Constraints:

(Hide) Options:

Hit: Context: Cluster: Tracking:

Searching for: "mealy machine is" "mealy machine refers to" "mealy machine means" "mealy machine will" "mealy machine helps" using: HotBot Infoseek AltaVista Excite Lycos Northern Light Yahoo WebCrawler

Tip: For better precision with multiple terms you might like to use "+" to ensure that the results contain specific terms (e.g. "free gifts" + "optics").

Ref...k, such that all state memory changes are made with respect to the clock signal. T F A Moore machine usually has less states than an equivalent Mealy machine. T F A potential problem with a Mealy Machine is that the output changes are not synchronized with clock changes. Fill in the blanks. 10 points at 2 points per blank The canonical SOP form of an expression results in a level circuit. ...

Ref... input alphabet and by creating multiple input mechanisms for reading events. Second, the transition function must be modified so that controller tasks can be performed during state transitions. A Mealy machine is a DFA that defines symbols which are output during state transitions. For the current purpose, a similar mechanism is used to perform controller tasks such as moving a robot, opening a vice of fi...

Ref... these general premises to the Collatz conjecture, which of all open problems at the moment is perhaps the most conveniently conducive to the approach. = Generalized Sequential Machines, GSMs A Mealy machine is a Finite State Automaton with a single output symbol associated with each state transition (e.g. see [1, p.42]). A GSM or Generalized Sequential Machine is similar to it is a FSA with an output string...

Ref... next state which then effect the output. (State refers to all latched events and values.) Argument for Mealy is that the output depends on the transition, thus ignoring the buffers, the CFSM is a Mealy machine. (Will explore this more later.) Issues concerning composition have not been resolved by the Pollis group, there is no composition as it stands. Resources A Formal Methodology for Hardware/Software...

Ref... State Machines We consider two types of state machines, Moore and Mealy. A Moore machine is a Mealy machine whose output does not directly depend on its input. Mealy Machines A Mealy Machine is a 6-tuple $M = (S, D, Q, q_0, D(q, q), l(q, q))$ where $S \neq \emptyset$ is a finite set of input symbols (we will use a to denote a particular input symbol) $D \neq \emptyset$ is a finite set...

Ref...ving on state register flip-flops, it is still desirable to use them. This leads to alternative synchronous design styles for Mealy machines. Simply stated, the way to construct a synchronous Mealy machine is to break the direct connection between inputs and outputs by introducing storage elements. One way to do this is to synchronize the Mealy machine outputs with output flip-flops. See Figure 8...

Ref...itions. A FSA is called non-deterministic if there is one or more transitions from one state to another for a given input. A Moore machine is an FSA which associates an output with each state and a Mealy machine is an FSA which associates an output with each transition. The Moore and Mealy FSAs are important in applications of FSAs. Equivalence of deterministic and non-deterministic fa it might seem ...

Ref...icle, and you, will make use of this three-block model to describe a state machine in VHDL using our four-step design procedure. Moreover, the outputs of a state machine define its type. That is, a Mealy machine is one in which the outputs are a function of both the inputs and the current state-variables (Figure 1). A Moore machine has outputs that are a function of the state-variables only (Figure 2). And a ML...

Ref... 60 Bestudeer van module 1.1 de blz.7 t/m blz.13 grondig. 20 Als u niet (of niet meer zo goed) weet wat een toestandsmachine, een toestandsdiagram, een Moore- of een Mealy-machine is, zoek dan wat u niet (meer) weet op in uw boek(en) over digitale techniek. 30. Maak de oefenopdracht van blz.13 10 Lees de rest van module 1.1 oppervlakkig ...

[...section deleted...]

【図27】

102

Home Options Help Suggestions Problems NECI meta search (13674 queries)

Find:

Web: ☐ Usenet: ☐ FTP: ☐ Images: ☐ CDD: ☐ Journals: ☐ Tech: ☐ All: ☐

(Show) Constraints:

(Hide) Options:

Hits: Context: Cluster: Tracking:

Tip: Clicking on the search engine links in the "Searching for:" line will show the search engine response to the current query.

Recently modified URLs:

Page: <http://ftp.kernel.org/pub/linux/kernel/testing/> [Stop tracking URL]

Recent documents matching: [signafy](#) [Mark as seen] [Stop tracking query]

INFORMIX-Universal Server Powers Comprehensive Media Asset Management Package From Virage... NW n/a 12k
<http://www.infoseek.com/Content?am=ix.BWR1674-19970915&q=signafy&col=NX&n=25&kt=A&ak=afnews>
 ... Group The Content Group, Excalibur Technologies Corp. Muscle Fish, LLC Silicon Graphics, Inc. Signafy, Inc.
 TECHMATH GmbH and TATA Consultancy Services. About INFORMIX-Universal Server INFORMIX-Un...

INFORMIX-Universal Server Powers Comprehensive Media Asset Management II n/a 13k
<http://www.infoseek.com/Content?am=ix.BWR1997091521525178X&q=signafy&col=IX&n=25&kt=A&ak=industrynews>
 ...roup The Content Group, Excalibur Technologies Corp. Muscle Fish, LLC Silicon Graphics, Inc. Signafy, Inc.
 TECHMATH GmbH and TATA Consultancy Services. About INFORMIX-Universal Server I...

Today at NECI

9:00 ... NJ Prog. Language Workshop - Multipurpose Rooms 2F00 2F01, 2nd Floor, (AW)

OUT ... Ebbesen, Gottlieb, de Ruyter, Thornbø

Recent articles about NEC Research in the press:

NEC Research Promises Terabit Memory Chips NT n/a 20k <http://www.techweb.com:80/wire/news/1997/09/0911nec.html>
 ... NEC Research Promises Terabit Memory Chips ... / ... NEC Research Promises Terabit Memory Chips ... / ... Chips
 International NEC Research Promises Terabit Memory Chips (09/11/97 12:00 p.m. EDT) By John Boyd, ...

[...section deleted...]

【図28】

<http://www.research.digital.com/SRC/publications/src-papers.html> New text.

1. Paul McJones and John DeTreville. Each to Each programmer's reference manual. Technical Note 1997-023, Digital Equipment Corporation Systems Research Center, Palo Alto, CA, October 1997.

2. Monika Henzinger and Han La Poutre. Certificates and fast algorithms for biconnectivity in fully-dynamic graphs. Technical Note 1997-021, Digital Equipment Corporation Systems Research Center, Palo Alto, CA, September 1997.

3. Monika Henzinger. Improved data structures for fully dynamic biconnectivity. Technical Note 1997-020, Digital Equipment Corporation Systems Research Center, Palo Alto, CA, September 1997.


4. Monika Henzinger and Valerie King. Maintaining minimum spanning trees in dynamic graphs. Technical Note 1997-019, Digital Equipment Corporation Systems Research Center, Palo Alto, CA, September 1997.

5. Marc Brown, Marc A. Najork, and Roope Raisamo. A Java-based implementation of Collaborative Active Textbooks. In *1997 IEEE Symposium on Visual Languages*, pages 372-379. IEEE Computer Society, September 1997. (PDF). (PostScript). (Copyright 1997 IEEE).

[...section deleted...]

Digital Systems Research Center

130 Lyton Avenue, Palo Alto, CA 94301
Tel: (415) 853-2100 Fax: (415) 853-2104



Legal notice
Send comments to the owner of this page.
Last modified: Tuesday, 07-Oct-97 10:32:46 PDT

Copyright Digital Equipment Corporation 1995-1997. All Rights Reserved.

フロントページの続き

(51)Int. Cl. ⁷	識別記号	F I	キーワード (参考)
G 0 6 F 3/00	6 5 4	G 0 6 F 3/00	6 5 4 D

(72)発明者	ステファン ロバート ローレンス	F ターム (参考)	5B075 ND06 ND36 NK02 NR12 NR14
	アメリカ合衆国、 ニュージャージー		PP30 PQ02 PQ22 PQ46 PQ48
	08540、 プリンストン、 インディペン		PR06 QM08
	デンス ウェイ 4 エヌ・イー・シー・		5E501 AA02 AB15 AC20 AC33 BA03
	リサーチ・インスティテューテュ・インク		BA09 CA02 FA13 FB04 FB28
	内		FB44